LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN
aprile 1973
numero 56
anno VI
volume X

L'economia dell'idrogeno

Un esempio di regime energetico in cui tutte le fonti di energia sono usate per produrre idrogeno da distribuire come combustibile non inquinante da impiegare per vari usi

di Derek P. Gregory

Il dilemma fondamentale rappresentato dalla cosiddetta « crisi mondiale dell'energia » può essere cosí formulato: mentre l'economia del mondo in generale e quella delle zone industriali in particolare dipendono sempre piú dal consumo di energia, aumenta la consapevolezza che le principali fonti di tale energia, ovvero le riserve non reintegrabili di combustibili fossili della Terra, sono destinate con assoluta certezza a esaurirsi e che comunque l'ambiente naturale non può assorbire con facilità i sottoprodotti dei consumi di combustibili fossili in quantità molto maggiori di quelle attuali senza essere sottoposto a livelli inaccettabili di inquinamento.

Di solito non si tiene conto del fatto che l'eventuale soluzione del problema energetico dipende non solo dallo sviluppo di fonti sostitutive di energia, ma anche dall'ideazione di nuovi metodi di conversione dell'energia stessa. Dopo tutto vi è ovunque abbondanza di energia « naturale », ma questa non si presenta sotto una forma adatta all'uso immediato e non si trova in una posizione abbastanza vicina al luogo di consumo. Negli USA la maggior parte delle ricerche e degli sviluppi in corso per risolvere il problema energetico è dedicata a realizzare nuovi sistemi di conversione in energia elettrica dell'energia chimica (derivata da combustibili fossili), nucleare (derivata da reazioni di fissione o di fusione) e solare (derivata direttamente dal Sole).

Gli impianti di fissione nucleare forniscono oggi circa l'1,6 per cento dell'elettricità consumata negli USA (per il resto sopperiscono gli impianti a combustibile fossile per l'82 per cento circa e gli impianti idroelettrici per il rimanente 16 per cento). Nell'ipotesi che sia possibile sviluppare reattori autofertilizzanti sufficientemente economici, si potrà fra poco eliminare qualsiasi preoccupazione a breve termine sulle limitate risorse di energia nucleare, e quindi gli impianti termonucleari potranno fornire per l'anno 2000 fino alla metà del fabbisogno americano di elettricità.

Se questa previsione è esatta e se il gap di energia del futuro sarà colmato dall'energia nucleare messa a disposizione dell'utente sotto forma di elettricità, gli USA si saranno molto avvicinati a una « economia interamente elettrica ». È già possibile intravedere questa tendenza osservando il grafico nella pagina seguente che mostra come negli USA la richiesta di elettricità sia attualmente in aumento a un ritmo molto superiore della richiesta complessiva di energia. Secondo alcune stime, mentre per l'anno 2000 il consumo globale di energia negli USA sarà raddoppiato, la richiesta di elettricità aumenterà di circa otto volte portando la quota dell'energia elettrica da circa il 10 a piú del 40 per cento del consumo totale di energia.

È naturale domandarsi quanto sia desiderabile questa tendenza verso una economia prevalentemente elettrica e anche, con maggior precisione, se ci siano altre forme di energia che possano essere erogate al punto di impiego in modo piú economico e meno visibile dell'energia elettrica.

Si prendano per esempio in considerazione le principali destinazioni della energia, cioè i trasporti, il riscaldamento e i processi dell'industria pesante, categorie di utenti che, soprattutto per ragioni di economia e di facilità di trasporto, vengono oggi alimentate essenzialmente con energia prodotta da combustibili fossili; questi, con l'esaurimento delle scorte, diventeranno piú costosi rendendo relativamente piú economico l'uso dell'energia elettrica di origine nucleare. In questa situazione si può fare un esempio di utilizzazione indiretta delle fonti di energia nucleare per la produzione di un combustibile sintetico secondario che sarebbe erogato in modo piú economico e usato in modo piú facile dell'elettricità per molti impieghi di grandi proporzioni. In questo articolo saranno esaminati i meriti di quello che l'autore considera il principale candidato per tale combustibile secondario: l'idrogeno gassoso.

Sotto molti aspetti l'idrogeno è il combustibile ideale. Sebbene non sia un combustibile « naturale », può essere facilmente ottenuto per sintesi dal carbone, dal petrolio o dal metano; ma piú importante è la possibilità di poterlo produrre dissociando semplice-

mente le molecole dell'acqua per mezzo di energia elettrica ricavata da una fonte di energia, per esempio da un reattore nucleare. Forse il piú grande vantaggio dell'idrogeno quale combustibile, almeno da un punto di vista ambientale, è che, quando brucia, l'unico prodotto di combustione è l'acqua! Nessuno dei tradizionali inquinanti del combustibile fossile – l'ossido di carbonio (CO), l'anidride carbonica (CO_2), l'anidride solforosa (SO_2), gli idrocarburi, le particelle, gli ossidanti fotochimici, eccetera – possono essere generati da una fiamma di idrogeno e l'esigua quantità di ossido di azoto (NO), formato dall'aria che partecipa alla combustione, può essere tenuta sotto controllo. Inoltre, poiché per principio le scelte delle sorgenti di energia sono circoscritte alle sostanze effettivamente « illimitate » come l'aria e l'acqua, l'idrogeno è di gran lunga il combustibile piú facilmente sintetizzabile.

In teoria si può quindi pensare a una economia dell'energia nella quale l'idrogeno venga ricavato dall'acqua con l'ausilio dell'energia elettrica, immagazzinato fino a quando non sia richiesto, trasmesso al punto di erogazione e ivi bruciato come combustibile per produrre elettricità, calore o energia meccanica (*si veda l'illustrazione nella pagina a fronte*). Un simile ipotetico modello non è privo di problemi e di svantaggi, ma a conti fatti i benefici sembrano essere tanto grandi da poter ritenere che, come esiste una tendenza verso una « economia elettrica », vi debba essere nello stesso tempo anche una tendenza verso una « economia dell'idrogeno ».

Proprio come l'industria alimentare e delle bevande ha trovato antieconomico raccogliere e riutilizzare i recipienti vuoti, cosí l'attuale industria dell'energia non può permettersi di raccogliere e rimettere in circolazione i « recipienti vuoti di energia », intendendo con tale locuzione i sottoprodotti della combustione necessaria per produrre l'energia medesima. In entrambi i casi il sistema di non richiedere depositi nè resa ha l'inconveniente di addossare all'ambiente l'onere del recupero e della riutilizzazione. A parte gli evidenti e nocivi effetti sull'atmosfera terrestre, questo tipo di ciclo energetico presenta l'ulteriore svantaggio di richiedere tempi estremamente lunghi, della durata di parecchi milioni di anni, per la formazione di nuovi combustibili fossili dall'anidride carbonica atmosferica (*si veda la figura a pagina 12*). Questa è la ragione fondamentale che sta portando verso l'esaurimento delle riserve di combustibili fossili. Nel ciclo dell'idrogeno invece, viene immessa nell'atmosfera solo acqua che si mescola rapidamente con quella abbondante e mobile esistente sulla crosta terrestre, mentre in un luogo differente altra acqua viene riconvertita in idrogeno. Il sistema è caratterizzato da un ritardo trascurabile e non disturba l'ambiente, da cui però dipende per la chiusura del ciclo, cioè per compiere la funzione di « vuoto di ritorno ». Nell'ipotesi di avere a disposizione un'abbondante provvista di energia nucleare o solare questo sistema può soddisfare la domanda con la massima elasticità senza esaurire alcuna risorsa naturale.

L'idea di usare l'idrogeno come combustibile sintetico è tutt'altro che nuova; Rudolf A. Erren, un inventore tedesco che lavorava in Inghilterra, propose nel 1933 di produrre idrogeno in grandi quantità impiegando l'energia elettrica disponibile nei momenti di scarso assorbimento. Erren aveva fatto uno studio approfondito per modificare i motori a combustione interna in modo da poterli alimentare con idrogeno; lo scopo principale della sua proposta era di eliminare l'inquinamento dovuto al gas di scarico delle automobili e di diminuire l'importazione in Inghilterra di petrolio grezzo. (È interessante osservare che 40 anni dopo gli USA e tutti i paesi industrializzati si trovano di fronte agli stessi problemi: l'inquinamento causato dalle automobili e la necessità di importare petrolio grezzo in quantità sempre maggiori.)

Anche altri hanno proposto di usare l'idrogeno come combustibile o come mezzo per immagazzinare energia. F. T. Bacon, un pioniere nello sviluppo delle celle a combustibile in Inghilterra, fin dagli anni trenta ha sempre avuto come obiettivo finale lo sviluppo di un sistema per immagazzinare energia sotto forma di idrogeno utilizzando celle elettrolitiche reversibili a combustibile. Piú recentemente, durante gli anni sessanta, la US Atomic Energy Commission (AEC) ha finanziato una serie di studi su complessi nucleari-agricoli-industriali detti « nuplexes » (dall'inglese *nuclear-agricultural-industrial complexes*), che traggono tutta l'energia loro occorrente da un unico reattore

La crescente tendenza verso un'economia « interamente elettrica » risulta evidente da questo grafico il quale mostra che negli USA la richiesta di elettricità (*linea in basso*) aumenta con un ritmo assai piú veloce che non la richiesta complessiva di energia (*linea in alto*). Se tale tendenza continuerà, gli USA si avvieranno a un'economia che, verso il XXI secolo, risulterà essere prevalentemente elettrica. I dati riportati nel grafico sono stati forniti dall'US Department of Commerce e dall'Edison Electric Institute.

In un'economia a idrogeno l'energia dovrebbe essere ottenuta da questo gas impiegandolo come combustibile secondario sintetico, dopo averlo ottenuto dall'acqua, in grandi centrali termonucleari o solari (*a sinistra*). L'idrogeno dovrebbe essere immesso in una rete di gasdotti interrati estesa all'intera nazione (*al centro*) che dovrebbe comprendere dei serbatoi per immagazzinare l'energia sotto forma di idrogeno gassoso in depositi sotterranei o di idrogeno liquido in cisterne in superficie. L'idrogeno verrebbe poi distribuito secondo le necessità degli utenti per essere adoperato sia direttamente come combustibile per riscaldamento, sia come materia prima per diversi processi chimici, sia come fonte di energia per la generazione locale di energia (*a destra*).

nucleare. Gli studi della AEC comprendevano il concetto di sottoporre l'acqua a elettrolisi per fornire l'idrogeno occorrente per la fabbricazione di fertilizzanti e di prodotti chimici. Negli ultimi due anni sono stati pubblicati diversi articoli in riviste scientifiche e tecniche che proponevano esaurienti studi sulla produzione, trasmissione, immagazzinamento e utilizzazione dell'idrogeno nei dispositivi di combustione e nei motori. Tali studi sono in corso negli USA e all'estero in alcune università e laboratori di ricerche industriali, compreso l'Istituto di tecnologia del gas di Chicago (ove opera l'autore), il cui lavoro è finanziato dall'American Gas Association.

La difficoltà di trasportare l'idrogeno ne ha sempre ostacolato l'impiego come combustibile. È chiaro che si devono trovare metodi migliori di quello della compressione in bombole di acciaio. Sono già in uso l'immagazzinamento e il trasporto dell'idrogeno allo stato liquido, come pure sono stati presi in considerazione con buone speranze gli idruri metallici e gli idruri sintetici organici e inorganici. Non esiste tuttavia nessuna ragione per non distribuire l'idrogeno nello stesso modo del metano e cioè con tubazioni interrate che raggiungono attualmente negli USA la maggior parte delle industrie e l'80 per cento delle abitazioni private.

Prima di valutare i meriti del concetto di economia dell'idrogeno è istruttivo considerare l'alternativa di un'economia interamente elettrica. Si supponga per un momento di non prendere in considerazione la sintesi di un combustibile chimico secondario; sarà allora necessario affrontare l'eventualità di generare e trasmettere enormi quantità di energia elettrica. Per soddisfare la crescente richiesta di elettricità sono già state costruite negli USA nuove centrali di dimensioni mai raggiunte prima. Alcuni anni fa una centrale termoelettrica da 500 megawatt era considerata gigantesca; oggi le centrali da 1000 megawatt sono abituali e sono già previsti per il futuro impianti con una potenza installata di 10 000 megawatt. Nonostante gli intensi sforzi dei progettisti, il rendimento delle centrali termoelettriche a vapore è abbastanza basso: circa il 40 per cento per un moderno impianto a combustibile fossile e il 33 per cento per un impianto nucleare (si veda l'articolo *La conversione dell'energia* di Claude M. Summers in « Le Scienze », n. 40, dicembre 1971). Da ciò deriva che questi grandi impianti, o gruppi di impianti, emettono notevoli quantità di calore di scarto e quindi devono essere installati vicini a grandi masse d'acqua che assicurino ampie possibilità di raffreddamento oppure in aperta campagna dove la dispersione del calore nell'atmosfera non dia luogo a effetti nocivi locali. Anche la preoccupazione sulla sicurezza dei reattori nucleari ha una forte influenza sulla ubicazione di tali impianti. È probabile che, a causa di queste limitazioni, le enormi centrali termoelettriche del futuro saranno costruite a una distanza di almeno 80 chilometri dai centri di consumo; sono già state addirittura studiate centrali situate su piattaforme galleggianti sul mare.

L'energia deve essere trasportata dai centri di produzione a quelli di consumo. Gli elettrodotti ad alta tensione su tralicci sono dispendiosi, per il costo sia delle attrezzature sia del terreno occupato e sono anche vulnerabili dagli agenti atmosferici. Inoltre l'industria elettrica sta incontrando in molte zone notevoli difficoltà per l'istallazione di nuovi elettrodotti aerei. I cavi sotterranei ad alta tensione di grande potenza costano almeno nove volte (e talora ben 20 volte) di piú delle linee aeree e quindi sono troppo costosi per un impiego su grandi distanze. I cavi interrati sono impiegati solo quando la spesa è giustificata da altre considerazioni, per esempio da motivi estetici o da una servitú di elettrodo eccessivamente costosa. Si stanno ora facendo molti studi per lo sviluppo di cavi criogenici superconduttori da interrare che permetterebbero il passaggio di elevate

densità di corrente a un costo ragionevole. Tuttavia attualmente questa tecnologia è ancora in una fase iniziale di sviluppo.

Sarebbe di grande utilità disporre di qualche sistema per immagazzinare la energia elettrica perché gli impianti di generazione lavorano con un rendimento maggiore quando funzionano a regime costante e a pieno carico. Siccome la richiesta degli utenti ha forti variazioni sia stagionali che giornaliere, occorre regolare continuamente la potenza generata. Il solo metodo oggi disponibile per l'immagazzinamento di grandi quantità di energia elettrica è costituito dagli impianti di pompaggio, ossia da centrali idroelettriche riversibili; purtroppo a sistemi di questo tipo si presta, da un punto di vista geografico, solo un numero limitato di luoghi.

Si vede quindi che alcuni dei problemi di fronte a cui si trova l'industria elettrica (la scelta delle località per le centrali termoelettriche, la spesa della trasmissione via cavo interrato e le scarse possibilità di immagazzinamento) sono resi piú complessi dai fattori che portano a realizzare centrali piú grandi e piú distanti. L'idea dell'economia a idrogeno potrebbe aiutare ad attenuare questi problemi.

L'idrogeno può essere trasportato e distribuito mediante gasdotti quasi nello stesso modo del metano. Il trasporto di combustibile in tubature è uno dei metodi piú economici di trasmissione dell'energia e l'idrogeno non potrebbe costituire un'eccezione. Di solito un sistema di distribuzione del gas è interrato e, oltre a non essere in vista, ha un minor ingombro di un elettrodotto. L'idrogeno può anche essere immagazzinato in grandissime quantità con l'identica tecnica usata oggi per il metano.

Prendiamo ora in considerazione la rete di trasmissione del gas esistente negli USA: nel 1970 erano in funzione gasdotti per un totale di oltre 400 000 chilometri che hanno trasportato nel corso dell'anno ben 634 miliardi di metri cubi di gas. Un tale sistema di condutture è necessario perché le fonti del gas metano si trovano in certe località del paese, mentre gli sbocchi di consumo sono in altre zone.

Nell'economia a idrogeno, l'idrogeno gassoso sarebbe prodotto da grandi impianti di energia nucleare (o solare) installati in località che presentino condizioni ideali per il raffreddamento nonché altre facilitazioni ambientali. Potrebbero essere inseriti in questa rete perfino generatori di idrogeno alimentati a carbone e situati vicino all'imboccatura delle miniere. Il sistema di trasmissione per mezzo di tubazioni verrebbe ulteriormente esteso per collegare queste località alle città in modo analogo al sistema di gasdotti per il metano.

La tecnologia per la costruzione e il funzionamento dei gasdotti per il metano è oggi completamente messa a punto. Un tipico gasdotto interurbano, lungo da 1000 a 1500 chilometri è composto da tubi di acciaio saldati aventi un diametro fino a 48 pollici (122 centimetri), interrati con un'adeguata protezione contro le avarie meccaniche e/o la corrosione elettrochimica. Il gas viene pompato da compressori disposti lungo la linea, generalmente a intervalli di 150 chilometri, e azionati con gas spillato dalla linea stessa. Di regola la pressione in linea è compresa fra 40 e 60 chilogrammi per centimetro quadrato, ma alcuni sistemi funzionano a oltre 70 chilogrammi per centimetro quadrato. Una tipica conduttura del diametro di 36 pollici (91 centimetri), ha una portata di 9,45 miliardi di grandi calorie all'ora o 11 000 megawatt nell'equivalente unità elettrica, pari approssimativamente a 10 volte la portata di una linea di trasmissione aerea monofase a 500 chilovolt.

Il metano non è il solo gas trasportabile in grandi gasdotti, quantunque nessun altro gas sia trasportato in tali quantità: l'anidride carbonica, l'ossido di carbonio, l'idrogeno e l'ossigeno possono essere tutti distribuiti attraverso gasdotti dello stesso tipo. L'industria finora non ha avuto alcun vantaggio economico che la spronasse a trasmettere idrogeno in forti quantità e a grandi distanze, ma quando trasporta idrogeno a brevi distanze impiega le pressioni e i materiali delle tubature convenzionali per il metano. Non c'è nessuna ragione tecnica che impedisca di trasportare l'idrogeno a qualsiasi distanza.

Poiché l'idrogeno ha un potere calorifico minore del metano (rispettivamente 2900 kcal/m^3 e 8900 kcal/m^3), per fornire la stessa quantità di energia ne deve essere trasportato un volume tre volte maggiore; la densità e la viscosità dell'idrogeno sono però tanto minori che lo stesso tubo può consentire una portata di questo gas tre volte maggiore con un leggero aumento dell'energia occorrente per i compressori. Quindi i gasdotti esistenti, se opportunamente ubicati, potrebbero essere impiegati per l'idrogeno con la medesima portata di energia.

Nell'economia a idrogeno sarà possibile immagazzinare quantità immense per compensare le variazioni giornaliere e stagionali del carico. Oggi il metano viene immagazzinato in due maniere: in « serbatoi » nel sottosuolo e come liquido criogenico. Negli USA il metano viene immagazzinato in 337 località nel cui sottosuolo esistono formazioni di rocce porose, con una capacità globale di 161 miliardi di metri cubi. La possibilità di immagazzinare l'idrogeno nelle rocce porose sotterranee sarà definitivamente accertata solo dopo che saranno stati eseguiti gli opportuni esperimenti. Comunque attualmente circa 850 milioni di metri cubi di elio, un gas avente bassa densità e caratteristiche di dispersione analoghe a quelle dell'idrogeno, sono immagazzinati con ottimi risultati in un serbatoio sotterraneo vicino a Amarillo nel Texas.

L'immagazzinamento criogenico del metano è una tecnica in rapido sviluppo: in 76 località degli USA sono in uso o in costruzione impianti per il « taglio dei picchi di carico » con energia ottenuta da metano liquefatto. Nessun motivo tecnico impedisce che un analogo procedimento di taglio dei picchi possa essere adoperato con l'idrogeno liquido. L'idrogeno liquido, inizialmente considerato una pericolosa curiosità di laboratorio, è già ora usato come un utile mezzo per immagazzinare e trasportare questo elemento a grandi distanze: esso viene normalmente spedito all'interno degli USA in carri cisterna ferroviari e in autocisterne. La tecnologia per la liquefazione dell'idrogeno e il riempimento dei serbatoi è già stata sviluppata soprattutto per l'industria spaziale. Il piú grande serbatoio di idrogeno liquido è infatti quello del Centro spaziale John F. Kennedy, che ha una capacità di 3,5 milioni di litri equivalente a 9,5 miliardi di grandi calorie o a 11 milioni di chilowattora (*si veda l'illustrazione in questa pagina*). Quantunque l'energia contenuta in questo serbatoio sia pari solo al 4 per cento circa dell'energia immagazzinata in un impianto per il taglio dei picchi a metano liquido, è anche pari al 73 per cento della capacità del piú grande invaso di pompaggio del mondo facente parte dell'impianto idroelettrico di Ludington nel Michigan.

La soluzione criogenica per l'immagazzinamento dell'energia ha il vantaggio di essere applicabile in qualsiasi posto, non avendo alcun vincolo né geografico né geologico, fattori che limitano sia l'immagazzinamento sotterraneo di gas sia gli invasi di pompaggio idroelettrici.

Il modo piú semplice di ottenere l'idrogeno mediante l'energia nucleare è l'elettrolisi, un processo nel quale una corrente continua viene fatta passare attraverso l'acqua, resa conduttrice con l'aggiunta di particolari sostanze, provocandone la decomposizione diretta nei suoi componenti elementari; l'idrogeno e l'ossigeno. I due gas vengono ottenuti completamente separati poiché si sviluppano ciascuno a uno dei due elettrodi. I sali o gli alcali, che devono essere aggiunti all'acqua per aumentarne la conducibilità, non vengono consumati e quindi la sola sostanza che deve essere aggiunta durante il processo è l'acqua pura.

Sono attualmente in funzione numerosi grandi impianti elettrolitici per produrre idrogeno nelle località dove questo elemento è richiesto (per esempio per la fabbricazione di ammoniaca e di fertilizzanti) e dove è disponibile

COMBUSTIBILI FOSSILI
IMMAGAZZINAMENTO E TRASPORTO
ENERGIA
COMBUSTIONE
OSSIGENO
AMBIENTE
ACQUA, ANIDRIDE CARBONICA, ECC.
VEGETAZIONE

Confronto fra gli effetti sull'ambiente dell'attuale ciclo energetico del combustibile fossile e di quello proposto per il combustibile idrogeno. Quando i combustibili fossili vengono bruciati per liberare la loro energia (*in alto*), l'ambiente deve ricevere i sottoprodotti della combustione. La formazione di nuovi combustibili fossili dall'anidride carbonica atmosferica richiede milioni di anni (*linea tratteggiata*). D'altra parte quando come combustibile viene bruciato l'idrogeno (*in basso*) il sottoprodotto della combustione è rappresentato dal solo vapor d'acqua che viene facilmente assimilato dall'ambiente. Il ciclo del combustibile si completa rapidamente senza esaurire risorse limitate né accumulare prodotti di scarico nocivi per l'ambiente circostante.

L'immagazzinamento dell'energia sotto forma di idrogeno liquefatto è già una procedura abituale dell'industria spaziale. Per esempio, questo serbatoio criogenico isolato sotto vuoto al Centro spaziale John F. Kennedy contiene 3,5 milioni di litri di idrogeno liquido per il rifornimento dei missili Apollo e può essere considerato il piú grande impianto di questo tipo attualmente esistente. In termini di energia l'idrogeno in esso contenuto è equivalente a 9,5 miliardi di grandi calorie o a 11 milioni di chilowattora.

energia elettrica a basso prezzo (di solito energia idroelettrica). Uno dei piú grandi impianti commerciali di elettrolisi del mondo è gestito dalla Cominco, nella Columbia Britannica (*si veda l'illustrazione a pagina 16*). Questo impianto consuma circa 90 megawattora e produce giornalmente circa 33 tonnellate di idrogeno per la sintesi dell'ammoniaca; l'ossigeno, ottenuto come sottoprodotto, viene usato in processi metallurgici. Esistono grandi impianti analoghi in Norvegia e in Egitto, e molti altri impianti piú piccoli per la produzione dell'idrogeno.

L'energia teorica richiesta per produrre idrogeno dall'acqua, è di 2,79 chilowattora per ogni metro cubo di idrogeno gassoso. In pratica il rendimento dei grandi impianti industriali è del 60 per cento circa e il consumo tipico di energia è di 5,3 chilowattora per metro cubo di idrogeno. L'energia consumata rappresenta la maggior parte del costo di esercizio dell'impianto: quindi, se si prende in considerazione l'impiego dell'idrogeno elettrolitico come combustibile, c'è un forte incentivo – in effetti una necessità – di incrementare il rendimento di questi impianti.

La cella a combustibile, soggetto di intense ricerche e sviluppi durante gli ultimi quindici anni nel quadro dei programmi spaziali, è in realtà una cella elettrolitica funzionante all'inverso. La cella a combustibile di piú semplice costruzione e funzionamento combina l'idrogeno e l'ossigeno, dando come prodotti acqua ed energia elettrica. Le celle a idrogeno-ossigeno sono state scelte e sviluppate per entrambi i programmi Gemini e Apollo grazie al loro alto rendimento che riduce la quantità di combustibile necessario a bordo per fornire al veicolo spaziale l'energia elettrica. Per sviluppare celle a combustibile ad alto rendimento sono stati necessari molti studi; la stessa tecnologia può essere sfruttata per accrescere il rendimento del processo inverso: la elettrolisi. (Le celle elettrolitiche in funzione nei laboratori aerospaziali hanno un rendimento superiore all'85 per cento.)

Aumentare il rendimento delle sole celle elettrolitiche ha relativamente scarsa importanza almeno finché il rendimento delle centrali termonucleari non supererà l'attuale basso valore, del 33 per cento circa. Tuttavia questa perdita di rendimento può essere impedita: per esempio Cesare Marchetti nei Laboratori italiani dell'Euratom, ha progettato un processo chimico per la scomposizione termica diretta dell'acqua in idrogeno e ossigeno usando il calore prodotto da un reattore nucleare. L'acqua, per essere scomposta direttamente nei suoi componenti, deve essere riscaldata a temperature molto elevate, circa 2500 °C, onde ottenerne la dissociazione. A tali temperature, che non sono disponibili nei reattori nucleari, i gas non possono essere separati in modo opportuno prima che si ricombinino. Si può però concepire una reazione in due fasi nella quale, per esempio, un metallo reagisca con il vapore a una temperatura ragionevole, per produrre idrogeno e un ossido del metallo. L'idrogeno viene separato con facilità dall'ossido che potrebbe a sua volta essere decomposto in ossigeno e nel metallo di partenza per riscaldamento. Sfortunatamente non sembra che esista nessun metallo in grado di sopportare una tale serie di reazioni a temperature abbastanza basse da essere compatibili con i reattori nucleari i cui materiali costruttivi limitano le temperature di funzionamento a circa 1000 °C.

L'idea di Marchetti è quindi una sequenza assai piú complessa di reazioni cui prendono parte bromuro di calcio ($CaBr_2$), acqua (H_2O) e mercurio (Hg), nelle quali, a eccezione dell'idrogeno e dell'ossigeno, tutti gli altri reattivi vengono riciclati. Ogni reazione avviene a temperature inferiori a 730 °C, raggiungibili in un reattore nucleare. Malgrado che il processo sembri fattibile, occorre ancora un certo lavoro di sviluppo per tentare di aumentarne il rendimento globale e abbassarne il costo fino a livelli accettabili.

Per un'economia a idrogeno la quantità di gas richiesta sarebbe immensa. Per esempio, se si dovesse oggi produrre l'idrogeno equivalente alla produzione globale di metano degli USA, si dovrebbe fornire in un anno una quantità di combustibile pari a 637 miliardi di metri cubi di gas ossia energia per $5,67 \times 10^{15}$ grandi calorie. Questo corrisponde a circa 1980 miliardi di metri cubi di idrogeno: se potessimo produrlo in impianti elettrolitici nucleari a un ritmo continuo per tutto l'anno sarebbe necessaria una quantità di energia superiore a un milione di megawatt. Negli USA l'attuale potenza elettrica installata è di 360 000 megawatt: ciò significa che si sta ipotizzando una quadruplicazione della elettricità generata con la conseguente necessità di costruire oltre 1000 nuove centrali elettriche ciascuna della potenza di 1000 megawatt, e questo in aggiunta alla sempre crescente richiesta di energia elettrica per altri usi. La potenza elettrica installata negli USA durante gli ultimi cinque anni è aumentata invece di « soli » 105 000 megawatt.

Comunque una simile formidabile impresa di incrementare la capacità di generazione dell'energia non deriva solo dalla scelta di un'economia basata sull'idrogeno. Siccome l'enorme consumo di combustibili fossili diminuirà negli anni futuri, si deve almeno trovare un'equivalente fonte sostitutiva di energia. Le cifre menzionate danno una idea del rivoluzionamento che si determinerà entro il prossimo cinquantennio nel campo dell'energia.

Attualmente l'idrogeno piú economico è ottenuto dal metano. Poiché ovviamente l'idrogeno non può essere meno costoso della sostanza da cui proviene, non si può pensare che possa sostituire come combustibile il metano. L'idrogeno elettrolitico è ancora piú costoso a meno che non sia disponibile energia elettrica a un prezzo molto basso. Gli attuali prezzi dell'elettricità si basano sull'alimentazione di un carico fluttuante ma la possibilità di immagazzinare l'idrogeno, eliminerebbe le punte di carico riducendo lievemente il prezzo dell'energia elettrica.

Il costo dell'idrogeno prodotto con elettricità deve necessariamente essere sempre superiore a quello dell'energia adoperata; il minor costo di trasmissione e di distribuzione dell'idrogeno lo rende però vantaggioso per gli utenti. Gli ultimi dati economici pubblicati dalle industrie elettriche e del gas consentono di dedurre le quote della produzione, della trasmissione e della distribuzione che gravano sui prezzi medi addebitati per il gas e l'elettricità a tutti i tipi di utenze. Questi dati possono a loro volta essere confrontati con quelli corrispondenti dell'idrogeno prodotto elettroliticamente (*si veda l'illustrazione a pag. 17*). Le cifre per l'idrogeno sono dedotte supponendo per ipotesi che tutta l'elettricità generata negli USA nel 1970 fosse stata convertita in idrogeno, e questo trasmesso attraverso l'esistente rete di trasmissione del metano (su una distanza media di 1500 chilometri) e distribuito agli utenti sotto forma di combustibile gassoso. L'onere dell'elettrolisi, pari a 2,2 dollari per milione di grandi calorie è stato ricavato dai preventivi dell'AEC per il costo di fabbricazione delle celle elettrolitiche di tecnica avanzata. I costi di trasmissione e di distribuzione dell'idrogeno sono basati sugli equivalenti costi del metano, modificati per tener conto delle proprietà fisiche e dei fattori di sicurezza differenti necessari per il trattamento dell'idrogeno.

Da questo confronto risultano evidenti due osservazioni: la prima è che oggi per l'utente medio è molto piú economico acquistare l'energia sotto forma di metano che non di elettricità; l'altra è che sarebbe già possibile vendere all'utente di gas energia sotto forma di idrogeno a un prezzo inferiore a quello pagato oggi per l'elettricità. Comunque è chiaro che tale idrogeno non troverà clienti finché il metano avrà i prezzi attuali.

Guardando al futuro si capisce che il prezzo del metano aumenterà rapidamente insieme con i prezzi di tutti i combustibili fossili; questi aumenti sono determinati dalle scarse riserve di combustibili, dall'influenza delle norme contro l'inquinamento e da fattori sociali come per esempio la tutela della natura e la salute dei lavoratori impiegati nell'industria mineraria. Invece per il prezzo dell'energia nucleare, sebbene sembri attualmente in rapido aumento, si può prevedere una certa stabilizzazione nell'era del reattore autofertilizzante in quanto allora non vi sarà nessuna notevole limitazione delle scorte.

Per il momento non è possibile pronosticare con precisione quale sarà il probabile costo dell'energia sotto forma di idrogeno, ma senza dubbio si può pensare che in futuro i prezzi di tutte le forme di energia saranno notevolmente piú elevati. Anche cosí, a lungo termine e al punto di erogazione, l'idrogeno sarà piú economico del metano e molto probabilmente anche dell'elettricità.

L'idrogeno, quando sarà disponibile dovunque come lo è oggi il metano, assolverà facilmente altre funzioni in aggiunta a quelle attuali del metano. L'idrogeno può essere adibito a usi domestici (cucina e riscaldamento) e industriali (riscaldamento); può servire inoltre come materia prima chimica per numerose industrie fra cui quelle petrolchimiche, metallurgiche, dei fertilizzanti e dei generi alimentari ed essere usato anche per generare elettricità in centrali locali.

Le proprietà di combustione dell'idrogeno sono assai diverse da quelle del metano: l'idrogeno brucia con una fiamma piú ferma e piú calda e le sue miscele con l'aria sono infiammabili entro limiti di composizione piú ampi. Questa diversità di caratteristiche fa sí che i bruciatori di idrogeno devono essere progettati in modo diverso da quelli per il metano e che, nella conversione da un gas all'altro, sarà indispensabile modificare tutti i bruciatori. Modifiche cosí estese non sono senza precedenti: un'analoga operazione fu effettuata quando gli USA sostituirono il gas di distillazione (contenente circa il 50 per cento di idrogeno) con il metano; alcune nazioni europee hanno di recente intrapresa la medesima conversione.

L'idrogeno, poiché non genera prodotti nocivi di combustione, può essere usato senza pericolo in installazioni prive di tiraggio. È quindi possibile concepire una camera di combustione domestica sprovvista di canna fumaria; in questo modo si risparmia il costo del camino e si aumenta del 30 per cento il rendimento del sistema di riscaldamento domestico alimentato a gas. Inoltre sono possibili cambiamenti piú radicali perché, tolta la necessità della canna fumaria, lo stesso concetto di riscaldamento centrale non ha piú motivo di sussistere: infatti ogni stanza può essere riscaldata con stufe singole a idrogeno senza dispositivo di scarico e indipendenti l'una dall'altra, mentre il vapor d'acqua che si forma determinerà un'utile umidità. Un altro cambiamento radicale è l'uso potenziale di riscaldatori catalitici. Poiché l'idrogeno è un carburante ideale per la combustione catalitica, è possibile realizzare un vero riscaldamento a gas « senza fiamma » mantenendo il letto catalitico a qualsiasi temperatura desiderata, anche a 100 °C: questa prospettiva promette di sconvolgere nel futuro i sistemi di riscaldamento domestico e le tecniche di cottura. A temperature cosí basse è praticamente impossibile che si producano ossidi di azoto, eliminando in tal modo l'unico eventuale inquinante del sistema a idrogeno.

L'idrogeno è anche il carburante ideale per le celle a combustibile: i problemi tecnologici che si sono presentati durante piú di un decennio nello sviluppo di celle a combustibile pratiche e commercialmente convenienti vengono semplificati moltissimo se si può usare l'idrogeno come combustibile. I generatori di elettricità costituiti da celle a combustibile funzionanti con idrogeno dovrebbero avere un rendimento non inferiore al 70 per cento: si può quindi realisticamente prevedere che essi possano trovare impieghi nelle abitazioni, negli edifici commerciali e industriali e nell'industria. Le centrali elettriche urbane di maggiori dimensioni potrebbero essere costituite da sistemi di celle a combustibile o da centrali a vapore alimentate a idrogeno. Il precedente concetto di un sistema di turbina a vapore a ciclo chiuso alimentato con idrogeno-ossigeno diventerebbe realizzabile mediante l'uso della tecnologia dei motori a razzo. Alcuni ricercatori dell'Istituto di Tecnologia del Massachusetts hanno proposto un sistema del genere per i sommergibili per il quale è stato comunicato che si può prevedere un rendimento complessivo del 55 per cento.

L'idrogeno è un ottimo carburante per i motori delle turbine a gas ed è stato proposto per essere usato negli aerei supersonici da trasporto. Per un impiego di questo genere è possibile immagazzinare e riempire i serbatoi con idrogeno liquido quale carburante. Anche se il grande volume necessario potrebbe rendere il suo uso meno interessante per gli aerei subsonici, la notevolissima economia di peso dà all'idrogeno un netto vantaggio in confronto a un equivalente carico di cherosene. Anche i motori convenzionali a combustione interna potranno funzionare a idrogeno, se verranno opportunamente modificati o riprogettati. R. J. Schoeppel dell'Università dell'Oklahoma e altri hanno dimostrato che, iniettando l'idrogeno nel motore attraverso una valvola in maniera analoga a come viene iniettato il carburante in un motore Diesel, si superano le caratteristiche di preaccensione dell'idrogeno. Altri, fra i quali Marc Newkirk della International Materials Corporation e Morris Klein della Pollution Free Power Corporation hanno riferito sul soddisfacente funzionamento dei motori convenzionali per automobili alimentati a idrogeno con modifiche al carburatore e ai collettori. Nel frattempo William J. D. Escher dell'Escher Technology Associates ha proposto un progetto radicalmente differente dei motori per automobili nel quale viene usato un sistema a vapore alimentato sia da idrogeno che da ossigeno. L'impiego dell'idrogeno liquido come carburante abituale per le automobili private è discutibile sul piano della sicurezza, ma è probabilmente applicabile alle aziende di trasporto pubblico che gestiscono linee di autobus e alle società di taxi.

Richard H. Wiswall jr. e James J. Reilly del Brookhaven National Laboratory hanno suggerito l'uso di idruri metallici per immagazzinare l'idrogeno destinato alla propulsione di veicoli. Un idruro di una lega di magnesio immagazzinerà energia sotto forma di idrogeno con la stessa efficacia (sulla base del peso) di un bombola di idrogeno liquido; però devono ancora essere superati alcuni problemi tecnici. Sembra che per ora non esista un modo semplice e chiaro per alimentare le automobili a idrogeno, ma sono in corso importanti studi per esaminare le varie scelte disponibili. In ogni caso se si dovesse sintetizzare un carburante liquido appropriato per le automobili e per gli aerei la sostanza di partenza deve essere l'idrogeno.

Una delle principali critiche al concetto di economia dell'idrogeno è che questo gas è troppo rischioso da usare. L'idrogeno è senza dubbio una sostanza pericolosa e deve essere trattata con le debite precauzioni; se però

viene maneggiato opportunamente con gli appositi dispositivi studiati per garantire la sicurezza, chiunque dovrebbe essere capace di usarlo senza rischi.

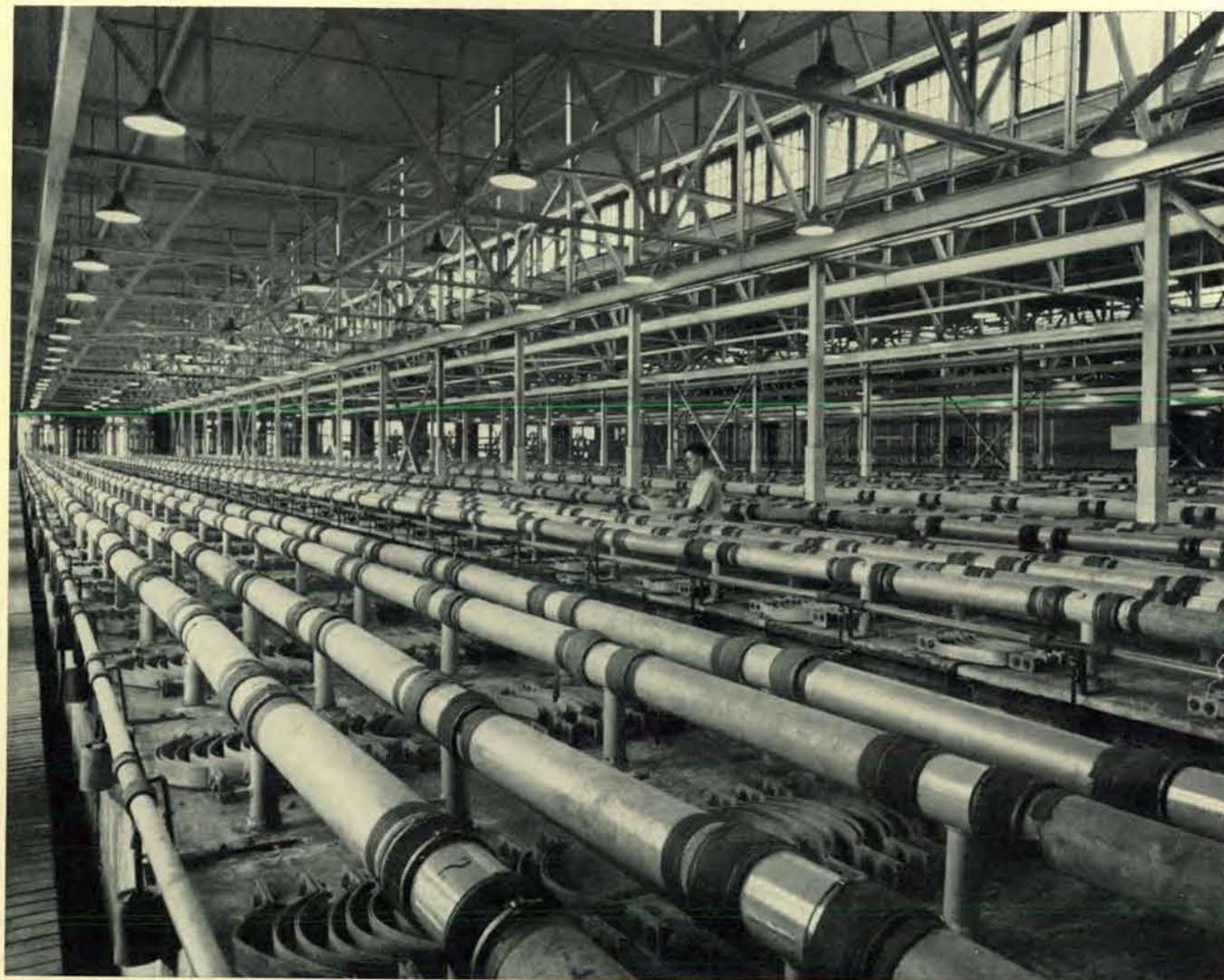
Quando si adoperava il gas di distillazione (ricavato dal carbone), che era composto per oltre il 50 per cento da idrogeno e che conteneva circa il 7 per cento di ossido di carbonio, si riusciva a vivere non solo con i pericoli di esplosioni e di incendi propri dell'idrogeno, ma anche con quelli dovuti alla tossicità dell'ossido di carbonio. Per mettere tutti in guardia contro un pericolo, è certo sufficiente un solo disastro di grandi proporzioni. Il piú famoso incidente causato dall'idrogeno, il disastro del dirigibile *Hindenburg* del 1937, è tuttora ricordato con sgomento. Difatti la paura quasi universale dell'idrogeno è stata definita la «sindrome dell'*Hindenburg* ». Per quanto spettacolare, quell'incendio si era pressoché esaurito nel giro di due minuti e delle 97 persone che si trovavano a bordo 62 sono sopravvissute.

Oggi sono imposte norme molto rigorose per l'uso del metano, perfino piú rigide di quelle applicate nell'industria per l'uso dell'idrogeno: la maggior parte di esse mirano realisticamente a ridurre le probabilità di infortuni. Proprio come siamo stati capaci di progettare dispositivi e procedimenti per consentire di riempire con la benzina i serbatoi delle nostre automobili e trasportare questa « bomba incendiaria » fino a velocità di oltre 100 chilometri all'ora su strade con traffico intenso e parcheggiarla di notte all'interno delle nostre case, saremo certamente in grado di escogitare norme di sicurezza per trattare l'idrogeno.

L'idrogeno non può essere percepito dai sensi: cosicché una fuga di idrogeno puro è estremamente pericolosa: tuttavia, per rendere manifeste le fughe di metano, si usano comunemente sostanze odoranti e lo stesso può essere fatto per l'idrogeno. Le sue fiamme sono quasi invisibili e perciò pericolose: per renderle visibili si potrebbe aggiungere al gas un illuminante.

I limiti di infiammabilità della miscela di idrogeno e aria sono oltremodo ampi: dal 4 al 75 per cento. È il limite inferiore, quasi uguale a quello del metano (5 per cento), che provoca il pericolo di incendio per una fuga di gas. Tuttavia l'idrogeno presenta il vantaggio di essere molto piú leggero dell'aria e di diffondersi a una velocità assai maggiore del metano: perciò una fuga di idrogeno potrebbe in realtà essere meno pericolosa di una fuga di metano. La proprietà pericolosa piú importante dell'idrogeno è che per incendiare la miscela infiammabile è sufficiente pochissima energia: solo un decimo di quella richiesta per accendere una miscela di benzina e aria o una di metano e aria e assai di meno del contenuto di energia di una scintilla prodotta da elettricità statica (probabile causa dell'incendio dell'*Hindenburg*, avvenuto proprio dopo un temporale con lampi e tuoni). Le norme di sicurezza saranno quindi fondate sull'ipotesi che se l'idrogeno è in condizioni di infiammarsi, ciò avverrà certamente. L'industria manipola enormi quantità di idrogeno senza correre rischi e senza infortuni proprio perché sono adottate le opportune precauzioni.

Per riassumere brevemente, le nostre riserve mondiali di combustibili fossili prima o poi si esauriranno e già si stanno avvertendo gli effetti del limitato approvvigionamento in quanto la energia elettrica prodotta con combustibile fossile è diventata piú cara. Occorre prepararsi entro i prossimi 50 anni a un notevole aumento di prezzo dell'energia prodotta da tutte le fonti e in modo particolare dal combustibile fossile. Un sistema di impiegare sia le fonti di energia nucleare sia le altre, è di usarle per trasformare in grandi impianti centralizzati l'acqua in idrogeno e adoperare poi quest'ultimo come un combustibile pulito e non inquinante. Questo è tecnicamente già possibile; devono essere realizzati solo alcuni perfezionamenti relativamente semplici, ben lontani dall'importanza degli studi tecnici necessari per sviluppare le fonti sostitutive di energia e cioè i reattori autofertilizzanti e i motori solari. La economia e la sicurezza sono i due ostacoli per lo sviluppo di questo sistema a idrogeno. Una combinazione dello sviluppo tecnico e l'attesa variazione dei prezzi relativi dell'energia può giustificare l'economia e opportune norme e progetti possono garantirne la sicurezza.

Se e quando ci metteremo in moto verso una economia a idrogeno il mondo presenterà certamente condizioni ambientali di gran lunga migliori di quelle attuali.

Questo grande impianto elettrolitico per la produzione di idrogeno dalla scomposizione dell'acqua è gestito dalla Cominco nella Columbia Britannica. Le 3200 celle elettrolitiche, che coprono piú di 8000 metri quadrati, consumano ogni giorno circa 90 megawattora per produrre circa 36 tonnellate di idrogeno che viene usato come prodotto fondamentale per la sintesi dell'ammoniaca.

In questo grafico i prezzi, al punto di erogazione, delle varie forme di energia sono stati scomposti negli oneri di produzione (*in colore scuro*), di trasmissione (*in colore intermedio*) e di distribuzione (*in colore tenue*). Il confronto mostra che, al momento attuale, è molto piú economico acquistare energia sotto forma di metano anziché di elettricità. Inoltre il grafico mostra che, sebbene il costo di produzione dell'idrogeno dall'elettricità debba essere sempre piú elevato del costo dell'elettricità stessa, i minori costi di trasmissione e di distribuzione consentono già di vendere agli utenti di metano energia sotto forma di idrogeno a un prezzo inferiore di quello attuale per l'elettricità.

Una veduta notturna dal ponte della *Glomar Challenger* che riprende la parte centrale della nave e la parte inferiore della torre di perforazione; alta 42 metri, quest'ultima è la struttura piú importante di questa particolarissima nave. Oltre alla torre di perforazione vi è una rastrelliera automatica in cui possono essere sistemate le aste per una lunghezza complessiva di 7200 metri. L'asta sospesa verticalmente al di sopra della piattaforma di perforazione è l'estremità superiore di una colonna di aste lunga 4500 metri, che, nel momento in cui veniva effettuata la fotografia, era stata montata allo scopo di perforare attraverso il fondale del Mediterraneo a circa 130 chilometri a occidente della Sardegna. Come la fotografia stessa testimonia, le operazioni di perforazione vengono condotte 24 ore al giorno. La *Glomar Challenger* è l'unica nave in grado di eseguire perforazioni nell'oceano aperto. Essa usa un sistema di posizionamento dinamico che le consente di rimanere in posizione pressoché stazionaria sopra la bocca del pozzo anche su un mare tempestoso. La nave, che è stata messa a punto dalla Global Marine Inc. e che è mantenuta operativa dalla Scripps Institution of Oceanography secondo un accordo con la National Science Foundation, prende il nome dalla prima vera nave oceanografica, la *Challenger*, che partí nel dicembre 1872.

Quando il Mediterraneo si disseccò

Durante la crociera che la nave da perforazione Glomar Challenger *ha compiuto in Mediterraneo nel 1970, sono emerse prove che sei milioni di anni fa il Mediterraneo era un bacino desertico profondo 3000 metri*

di Kenneth J. Hsü

Sei milioni di anni fa una vera e propria rivoluzione biologica si estese rapidamente attraverso tutto il Mediterraneo. L'antica fauna mediterranea, costituita da discendenti di razze miste originarie degli oceani Atlantico e Indiano, compí un disordinato esodo in massa alla ricerca di un rifugio a occidente di Gibilterra. Chi rimase si trovò presto a dover affrontare la distruzione totale, fatta eccezione per qualche specie particolarmente resistente che poteva sopportare il progressivo deteriorarsi dell'ambiente. In questo modo si concluse il Miocene, il periodo immediatamente precedente il Pliocene, cui seguí il Quaternario in cui viviamo. Con l'inizio del Pliocene, o in un momento ancora piú vicino a noi, i transfughi ritornarono, accompagnati da nuove specie provenienti dall'Atlantico. Essi erano gli antenati dell'attuale fauna marina del Mediterraneo. Questi eventi drammatici, testimoniati dalle associazioni fossili di certe rocce sedimentarie italiane, non sfuggirono all'attenzione di C. Lyell, uno dei fondatori della geologia. La fine di questa rivoluzione, segnalata dall'instaurarsi di una nuova dinastia faunistica, fu scelta da Lyell nel 1833 come l'evento storico che divide il Miocene dal Pliocene. Ma quale fu la causa di questa rivoluzione?

Verso la fine del XIX secolo fu scoperta, durante un ricerca d'acqua sotterranea, una profonda valle sepolta al di sotto della pianura di Valence, nella Francia meridionale. La valle risultava scavata in una dura roccia granitica fino a una profondità di centinaia di metri al di sotto del livello del mare. La valle era riempita di sedimenti pliocenici di tipo oceanico, a loro volta coperti dalle alluvioni deposte dal Rodano. Non appena la valle sepolta fu scoperta, si trovò che essa si estendeva per circa 24 chilometri fra Lione e Valence. Col passare del tempo, il tracciato della valle sepolta fu individuato per piú di 160 chilometri sotto corrente fino a La Camargue nel delta del Rodano, dove il fondovalle fu raggiunto da una perforazione alla profondità di circa 900 metri al di sotto della superficie. Ovviamente l'attuale Rodano è un fiume debole e pigro in confronto con quel suo progenitore che si scavò un reticolo idrografico di dimensioni confrontabili quasi con il Gran Canyon del Colorado. Cosa causò una cosí intensa attività erosiva del Rodano?

Nel 1961 la nave oceanografica americana *Chain* entrò in Mediterraneo con un nuovo tipo di apparecchiatura sismica, allo scopo di esplorarne i fondali. L'apparecchiatura, detta CSP, che consentiva di ottenere profili sismici continui, può essere considerata come una forma perfezionata di ecoscandaglio. Oltre a registrare i segnali acustici riflessi dalla superficie del fondo, lo strumento captava anche i segnali rinviati dalle onde sonore che erano in grado di penetrare al di sotto della superficie del fondo per essere poi riflesse da strati compatti a centinaia di metri di profondità. Il nuovo strumento permise una scoperta di grande interesse: si trovò che tutto il sottofondo del Mediterraneo è interessato da un gran numero di strutture a pilastro, ciascuna delle quali con un diametro di qualche chilometro e dell'altezza di centinaia di metri, che si spingono verso l'alto entro gli strati di sedimenti sovrastanti (*si veda l'illustrazione a pagina 22 in alto*).

Per i geofisici queste strutture non rappresentavano una novità; esse assomigliavano molto a duomi salini.

I duomi salini si formano per la risalita nei sedimenti sovrastanti di una roccia evaporitica; essi sono comuni, per esempio, lungo la Gulf Coast degli USA ove sono accompagnati da numerosi giacimenti di petrolio. Per i geologi non è una sorpresa trovare del sale in sedimenti costieri poiché essi hanno per lungo tempo pensato che le formazioni evaporitiche fossero il risultato della deposizione in ambiente lagunare costiero. Fu invece del tutto inaspettata la scoperta di duomi salini al di sotto delle pianure abissali del Mediterraneo. Da dove poteva venire quel sale? Oppure, quelle strutture erano veramente duomi salini?

L'esplorazione del Mediterraneo continuò negli anni 60. W.B.F. Ryan, del Lamont-Doherty Geological Observatory, che ha partecipato a quattro crociere nel Mediterraneo, e altri studiosi che si interessavano a questa regione, rimasero assai colpiti dalla presenza in tutto il Mediterraneo di un orizzonte riflettente delle onde acustiche chiaramente delineato (*si veda l'illustrazione a pagina 22 in basso*). Tale orizzonte, detto orizzonte *M*, si trova comunemente a poche centinaia di metri al di sotto della superficie del fondo, e la sua geometria segue abbastanza da vicino la morfologia del fondale stesso.

Uno strato che può restituire riflessioni distinte deve essere assai compatto. Ma i sedimenti oceanici sono comunemente costituiti da minuscoli scheletri di piccoli organismi chiamati foraminiferi e nannoplancton. Che cosa poteva essere quello strato compatto? Perché tale roccia compatta doveva trovarsi a quella profondità nel Mediterraneo?

Per risolvere questi e molti altri problemi, il gruppo di scienziati che costituiva il Mediterranean Advisory Panel del JOIDES (Joint Oceanographic

Il Mediterraneo nel Miocene superiore: un mare sovrasalato

Poiché lo schema del « disseccamento » del Mediterraneo, proposto in questo articolo da Kenneth J. Hsü, non raccoglie l'unanime consenso dei geologi che si occupano di tali problemi, la Direzione della rivista ha creduto opportuno pubblicare anche la seguente nota di Raimondo Selli, direttore del Laboratorio di geologia marina del CNR, che presenta una interpretazione alternativa.

L'ipotesi di un generale disseccamento del Mediterraneo durante il Miocene superiore, cioè in quell'intervallo di tempo che i geologi chiamano Messiniano, è indubbiamente molto suggestiva. Lo stesso dicasi della successiva catastrofica inondazione dell'inizio del Pliocene che avrebbe ristabilito il regime marino normale nel Mediterraneo. Però il nostro modello del Mediterraneo durante il Messiniano è diverso; è, diremmo cosí, un modello « oceanico » e non « desertico ».

A nostro avviso i risultati delle perforazioni JOIDES sono di importanza fondamentale, però dobbiamo tener presente anche tutte le conoscenze acquisite dalle altre ricerche di geologia marina e soprattutto quanto è noto da piú di un secolo di ricerche in campagna, in miniera e mediante pozzi nella terraferma. I terreni del Messiniano sono infatti diffusi non solo al disotto del mare, ma anche sul continente soprattutto in Italia.

In tutte le aree affioranti il Messiniano è caratterizzato da evaporiti, cioè da rocce che si separano per evaporazione dell'acqua marina. A mano a mano che l'acqua marina evapora, si ha dapprima la precipitazione del carbonato di calcio quando l'acqua raggiunge la salinità di circa il 60 ‰ (acqua marina normale 35 ‰); successivamente comincia a separarsi il solfato idrato di calcio (salinità 120 ‰), poi l'anidrite (salinità 250 ‰), poi il salgemma (salinità 420 ‰) e infine numerosi altri sali principalmente cloruri e solfati di magnesio, potassio e sodio.

La successione calcare, gesso e anidrite è comune a tutte le serie messiniane del Mediterraneo. In qualche area molto limitata della Toscana, Calabria e Sicilia e nel Mediterraneo occidentale (pozzo JOIDES 134) si ha anche salgemma; in qualche rarissimo punto della Sicilia anche cloruri e solfati di potassio e magnesio. Il salgemma e gli altri sali piú solubili si trovano però nella metà inferiore della successione evaporitica e non alla fine, come farebbe supporre l'ipotesi del disseccamento con successiva inondazione pliocenica.

Durante il Messiniano inoltre si ebbero circa 18-20 cicli evaporitici, ognuno con progressiva concentrazione delle acque fino alla precipitazione dell'anidrite. Solo nei rari casi suddetti si ebbe anche la precipitazione degli altri sali piú solubili.

Ogni ciclo evaporitico fu bruscamente interrotto da diluizioni, che portarono l'acqua del Mediterraneo non molto lontana dalla salinità normale, tanto da permettere il ripopolamento con faune molto selezionate, ma tipicamente marine e spesso di mare profondo e aperto (Foraminiferi, Radiolari, pesci batipelagici ecc.). Non si tratta solo di specie indigene sopravissute in qualche modo ai periodi di forte salinità, ma anche di nuove specie planctoniche atlantiche introdotte nel corso del Messiniano.

Durante gli intervalli di diluizione si depositarono solo argille, marne e sabbie; cioè materiali terrigeni portati a mare, spesso in enorme quantità, dagli antichi fiumi e torrenti. Le argille e le marne sottostanti e intercalate alle evaporiti sono quasi sempre laminate e ricche di idrocarburi (petrolio, gas e bitume); talora lo sono le evaporiti stesse. Si hanno cioè tutti i caratteri di un tipico ambiente eusinico (come quello dell'attuale Mar Nero), con acque superficiali mosse e ossigenate, dove vivono gli organismi, e acque di fondo stagnanti e avvelenate per acido solfidrico, dove la sostanza organica viene ridotta a idrocarburi. La produzione di sostanza organica e le successive trasformazioni devono essere state enormi, tanto che i sedimenti del Messiniano sono considerati come le rocce madri, da cui sono derivate la stragrande maggioranza dei giacimenti italiani di metano e petrolio.

Per esempio nelle evaporiti e nei sedimenti terrigeni interposti si intercalano spesso materiali caotici di frana sottomarina e torbiditi (cioè materiali trasportati in sospensione da correnti di densità su scarpate sottomarine per decine o centinaia di km), che in certi casi dimostrano profondità dell'antico mare anche superiori ai 1000 m. Del resto perché si possa instaurare un ambiente eusinico, sono necessarie profondità almeno di parecchie decine o qualche centinaio di m per permettere la formazione di acque profonde stagnanti avvelenate.

Un altro punto da prendere in considerazione è lo spessore dei sedimenti messiniani che, dove le successioni sono complete, variano da un centinaio a oltre 2500 m; spesso prevalgono le evaporiti, talora i sedimenti terrigeni, soprattutto le turbiditi (p. es. bacino adriatico). Sempre però sono presenti e tipiche le evaporiti con calcare, gesso e anidrite, che hanno spessori complessivi variabili da 10 m circa fino a oltre 300 m. Il salgemma e i sali piú solubili, da soli possono anche superare gli 800 m. Sono cifre enormi che richiedono l'evaporazione di una quantità enorme di acqua marina normale. Infatti se immaginassimo di sbarrare oggi gli stretti di Gibilterra e dei Dardanelli, in modo da disseccare il Mediterraneo per evaporazione, avremmo la precipitazione media di soli 77 cm di calcare, gesso e anidrite, e di 23,5 m di salgemma e sali piú solubili. Sicuramente poi il Mediterraneo messiniano era mediamente meno profondo di quello attuale. È perciò necessario ammettere che non solo durante gli intervalli di diluizione, ma anche durante quelli a forte concentrazione vi potesse essere una continua entrata di acqua oceanica nel Mediterraneo per consentire l'accumulo delle grandi quantità di evaporiti messiniane. Infatti per ottenere 30 m di calcare, gesso e anidrite (spessore medio certamente molto inferiore a quello reale) sarebbe necessario far evaporare una quantità d'acqua 38 volte maggiore di quella contenuta nell'attuale Mediterraneo.

A questo punto, per spiegare i fatti esposti, viene naturale l'interpretazione delle evaporiti mediterranee secondo un modello « oceanico » anziché « desertico »; esso è schematizzato nella figura. Tale modello « oceanico » fu già ipotizzato dallo scrivente circa 20 anni fa, quando ammise per la prima volta che le evaporiti messiniane dovevano estendersi su tutto il Mediterraneo occidentale e centrale.

Prima e durante il Messiniano non esisteva ancora lo stretto di Gibilterra e le comunicazioni con l'Atlantico dovevano avvenire mediante lo stretto nord-betico (attraverso la Spagna meridionale) e forse quello sud-rifano (attraverso il Marocco settentrionale). All'inizio del Messiniano, per il sollevamento delle catene della Sierra Nevada e del Rif i due stretti si ridussero enormemente in larghezza e profondità, tanto che la corrente di entrata dall'Oceano al Mediterraneo doveva prevalere di gran lunga su quella contraria di uscita, per riuscire a compensare le forti perdite per evaporazione del nostro mare (il clima era allora subtropicale, piú caldo di quello attuale). Si instaurava cosí l'ambiente eusinico e per il continuo afflusso di acqua oceanica aumentava progressivamente la salinità fino al 250-300 ‰ e fino alla precipitazione dell'anidrite. Temporanei abbassamenti della soglia (o innalzamenti eustatici del livello oceanico) permettevano piú ampi scambi con l'Atlantico, tanto da provocare la diluizione dell'acqua mediterranea fino forse al 50-60 ‰ e il ripopolamento delle acque superficiali e costiere da parte degli organismi. Tali oscillazioni si ripeterono per 18-20 volte determinando i 18-20 cicli evaporitici.

Ai margini di questo mare che potremmo chiamare « dell'anidrite » e a esso collegati con strette soglie, si avevano bacini secondari, nei quali le acque subivano talora un'ulteriore concentrazione fin oltre il 420 ‰, con separazione del salgemma e in qualche caso dei sali piú solubili.

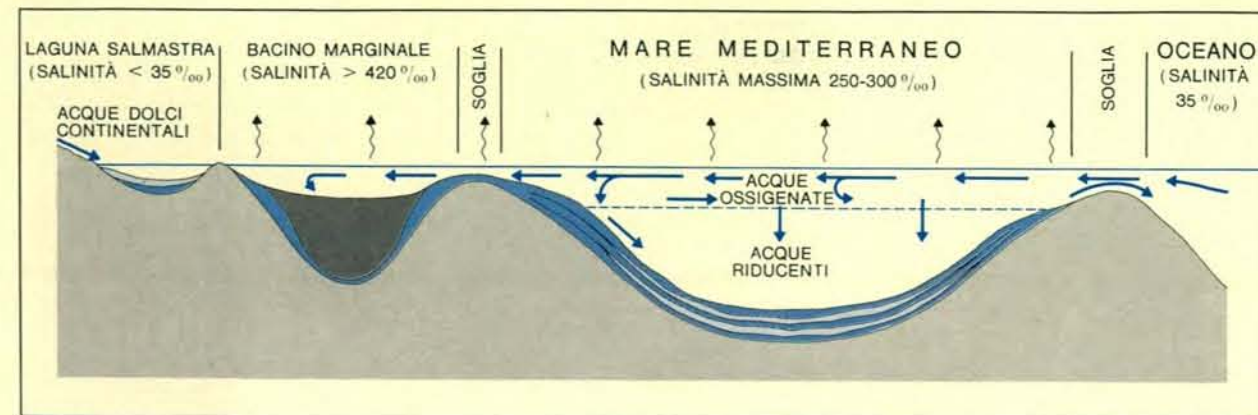
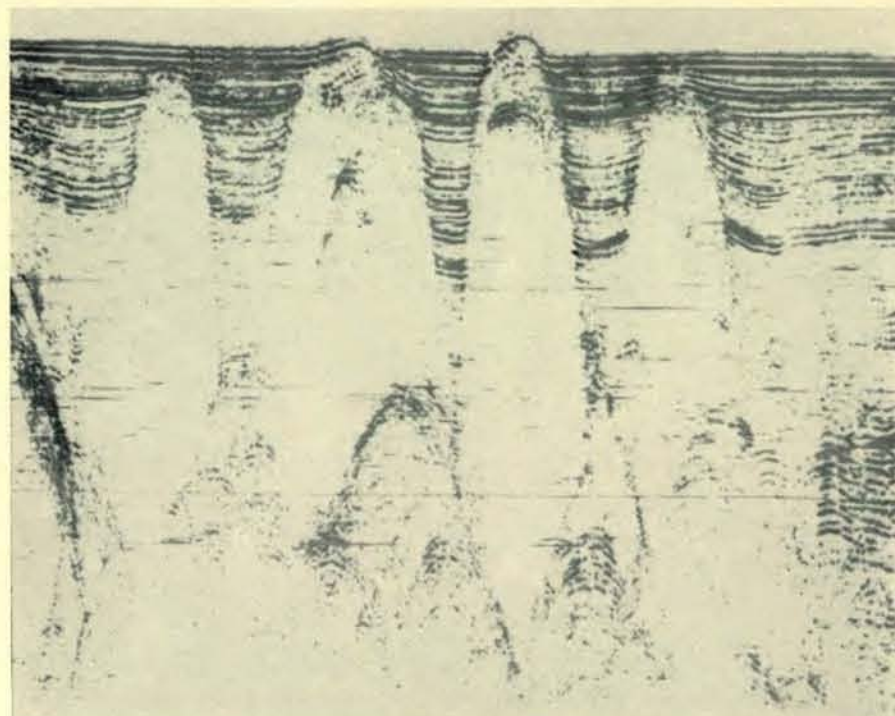
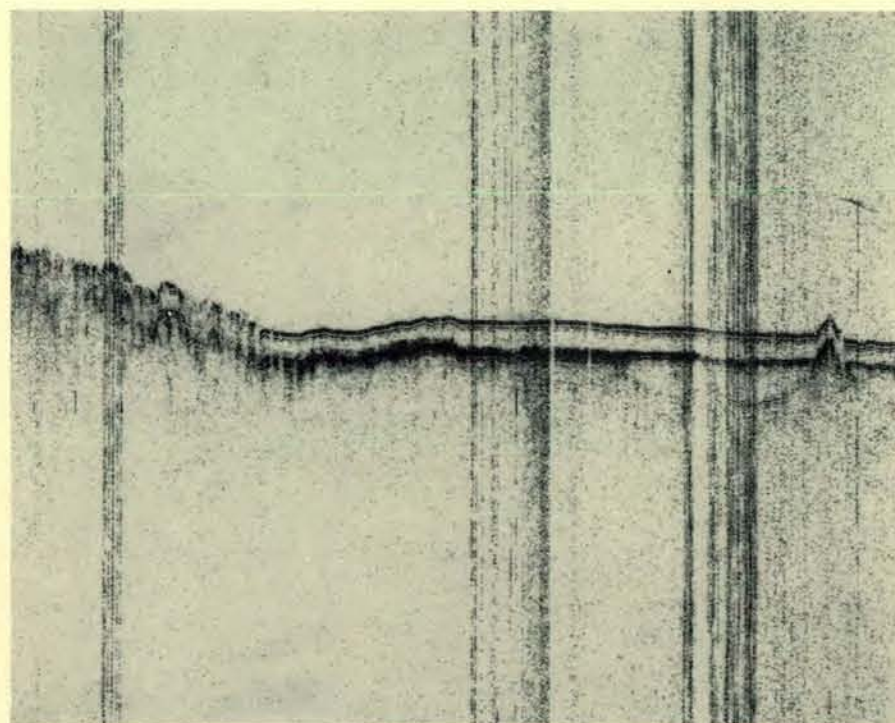
In prossimità delle coste dove sboccavano corsi d'acqua si potevano formare invece lagune o aree salmastre, dove nei periodi di diluizione delle acque mediterranee prosperavano faune dulcicole o salmastre, cioè di acque sottosalate (soprattutto Molluschi quali Bitinie, Cardi e Congerie).

Con questo modello « oceanico », cioè ammettendo la continuità di comunicazioni marine e il medesimo livello fra Mediterraneo e Atlantico, è possibile spiegare mediante oscillazioni della soglia o del livello marino tutti quei fatti (salinità e sue variazioni, spessore delle evaporiti, ambiente eusinico, faune marine intercalate alle evaporiti, ecc.), che altrimenti sarebbero difficilmente spiegabili con il modello « desertico ».

Del resto le condizioni del Mediterraneo durante il Messiniano rispecchiano in forma esagerata quelle attuali. Anche oggi infatti il nostro mare ha un enorme deficit idrologico; perde infatti per evaporazione in media circa 10 km³/giorno di acqua e solo 3,4 sono gli afflussi di acque continentali e le precipitazioni; il deficit di 6,6 km³/giorno gli è compensato dall'Atlantico attraverso Gibilterra. Basterebbe un restringimento di questo stretto per riportarlo a condizioni evaporitiche simili a quelle del Messiniano.

Con l'inizio del Pliocene si aprí lo stretto di Gibilterra. Le acque sovrasalate mediterranee si mescolarono rapidamente con quelle oceaniche fino a portarsi all'incirca alla salinità attuale (38-39 ‰); faune marine normali ripopolarono il Mediterraneo.

Indubbiamente i terreni del Pliocene piú antico in Italia e altrove sono spesso trasgressivi, discordanti e accompagnati da un conglomerato di base; cioè si sono depositati in seguito a una avanzata del mare su terreni del Messiniano o piú antichi, dopo che questi erano stati piú o meno a lungo emersi ed erosi. Ma pensiamo che questo sia dovuto ad abbassamento delle aree interessate e non costituisca una prova della « catastrofica inondazione » supposta dal prof. Hsü, tanto piú che mancano ovunque quei depositi desertici che l'avrebbero preceduta. In ogni modo è certo che nel bacino adriatico, dai dintorni di Bologna al Gargano, in certe zone della Toscana, Calabria, Sicilia e altrove vi è stata perfetta continuità di regime marino attraverso tutto il Messiniano e Pliocene, malgrado i profondi cambiamenti di salsedine.

In conclusione il modello « desertico » è indubbiamente interessante; ma mi sembra che vi siano molti argomenti per dubitare della sua validità. (*R.S.*)

Durante il Messiniano il Mediterraneo comunicava con l'oceano mediante una soglia molto piú ridotta di quella attuale di Gibilterra (*a destra*); per l'elevata evaporazione (*frecce in nero*) prevaleva la corrente di entrata; di qui il progressivo aumento di salinità fino alla precipitazione dell'anidrite (piú intensa alle basse profondità, minore sul fondo del bacino). Quando aumentava la profondità della soglia, le acque si diluivano, aumentavano gli apporti terrigeni, le torbiditi e le frane sottomarine, riprendeva la vita. A sinistra un bacino secondario marginale ultrasalato con la precipitazione del salgemma. All'estrema sinistra una laguna salmastra. In nero sono mostrati il salgemma e i sali piú solubili; in colore intenso calcare, gesso e anidrite; in colore chiaro i sedimenti terrigeni turbiditici e le frane sottomarine.

Strutture a pilastro, interpretate come duomi salini, in un profilo sismico continuo tratto da una sezione della lunghezza di circa 16 chilometri tracciata attraverso il piano abissale balearico nel Mediterraneo occidentale. L'apparecchiatura sismica registra non soltanto le riflessioni restituite dalla superficie del fondale ma anche quelle provenienti dalle onde sonore in grado di penetrare all'interno delle rocce che lo costituiscono per essere riflesse soltanto dai livelli piú compatti. Alcuni dei duomi visibili in questa immagine sono completamente sepolti; altri, invece, si spingono al di sopra della superficie del fondale a formare piccoli rilievi. La scoperta di queste strutture costituí la prima indicazione di estesi depositi salini sotto dei fondali del Mediterraneo.

L'orizzonte *M*, un orizzonte fortemente riflettente che si estende al di sotto di gran parte dei fondali del Mediterraneo. L'andamento di questo orizzonte (*tracciato inferiore*) rispecchia abbastanza fedelmente la topografia del fondale (*tracciato superiore*). La perforazione ha mostrato che l'orizzonte *M* corrisponde alla superficie superiore di una estesa formazione evaporitica sepolta, costituita dai depositi precipitati dalle acque soprasalate del Mediterraneo che sono andate formandosi quando, circa 6 milioni di anni fa, esso rimase isolato dall'oceano Atlantico. Il profilo è stato ottenuto dalla *Glomar Challenger* in rotta, da sud est della Sicilia (*a destra*), verso Creta (*a sinistra*).

Institutions for Deep Earth Sampling Program) nel 1969 raccomandò al Deep Sea Drilling Project (programma di perforazioni in mare profondo, sovvenzionato dalla National Science Foundation e amministrato dalla Scripps Institution of Oceanography degli USA) che la nave oceanografica per perforazioni in alto mare *Glomar Challenger* fosse inviata in Mediterraneo. La proposta fu approvata e la crociera di due mesi compiuta dalla *Glomar Challenger* nel Mediterraneo nell'autunno del 1970 (si vedano anche gli articoli di M. B. Cita Sironi, *Deep Sea Drilling Project: primi risultati*, in « Le Scienze » n. 16, 1969 e di C. F. Wezel, *I carotaggi profondi nel Mediterraneo*, in « Le Scienze » n. 47, 1972) forní risposte sorprendenti.

La rivoluzione biologica del Mediterraneo, l'incisione profonda e sepolta del Rodano e lo strato oceanico di sale sono tutte testimonianze silenziose del fatto che in un momento attorno ai sei milioni di anni fa il Mediterraneo era quasi completamente asciutto.

Lo strato compatto che costituisce l'orizzonte *M* è formato dai resti inorganici lasciati dal disseccamento del Mediterraneo e cioè da quelle particolarissime rocce che sono note con il nome di evaporiti.

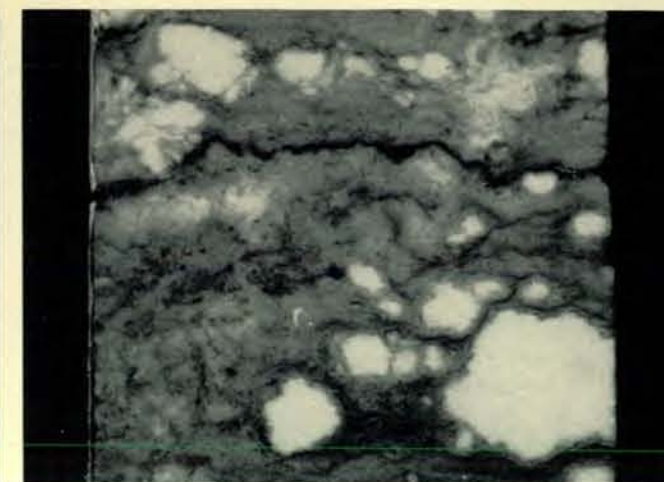
La *Glomar Challenger* lasciò Lisbona il 13 agosto 1970 per affrontare la 13ª crociera del Deep Sea Drilling Project. La nave è unica nel suo genere a causa della capacità che essa possiede di rimanere in posizione pressoché stazionaria anche su un mare in tempesta.

Guidati da un sistema di radiofari e di calcolatori, i quattro propulsori della nave possono mantenerla in posizione al di sopra della bocca del pozzo sul fondale marino entro un cerchio che ha un diametro complessivo di circa 30 metri.

Ryan e io eravamo gli scienziati responsabili di questa crociera e guidavamo un gruppo internazionale di 20 scienziati e tecnici. La sera del 23 agosto arrivammo in un punto che si trovava all'incirca 160 chilometri a sud est di Barcellona.

Dopo aver messo in posizione la nave al di sopra di un vulcano sottomarino sepolto, eravamo pronti per affrontare uno dei problemi piú interessanti e cioè quello dell'identificazione dell'orizzonte *M*.

La colonna di aste di perforazione fu fatta scendere fin sul fondo del mare che in quel punto era profondo circa 1800 metri; furono perforati circa 180 metri di sedimenti. Secondo i dati sismici in nostro possesso, avremmo dovu-

Laminazione convessa verso l'alto di una roccia sedimentaria estratta, tramite una perforazione eseguita dalla *Glomar Challenger*, dalla formazione la cui superficie costituisce l'orizzonte *M* del Mediterraneo; questa struttura è stata riconosciuta come una stromatolite. Le stromatoliti sono strutture tipiche delle rocce carbonatiche; esse devono la loro origine al fatto che, su certe pianure costiere, le alghe blu-verdi tendono a costituire sottili tappeti in cui vengono intrappolati granuli calcarei portati dal mare durante le periodiche sommersioni dovute alle tempeste. Dato che l'esistenza delle alghe dipende dalla fotosintesi, la presenza di una struttura stromatolitica è considerata una prova di deposizione in acque poco profonde; i fondali piú profondi, infatti, sono troppo scuri per i processi fotosintetici.

Noduli di colore chiaro, costituiti da anidrite, che compaiono con evidenza in questo campione di roccia estratto dai fondali del Mediterraneo; l'anidrite è la forma di alta temperatura del solfato di calcio e può precipitare da acque soprasalate che si trovino a temperature superiori ai 35 °C. A temperature inferiori, infatti, precipita il gesso che è la forma di bassa temperatura del solfato di calcio. L'anidrite attualmente si forma quasi esclusivamente nei depositi sedimentari di quell'ambiente particolare dei deserti caldi e costieri aridi chiamato sebkha; qui l'anidrite si deposita nelle porosità dei sedimenti in prossimità della superficie superiore della falda acquifera. L'anidrite si accresce sotto forma di noduli irregolari tipici che hanno un aspetto analogo alle concrezioni che accade di trovare nei suoli aridi.

to ottenere il risultato che ci si proponeva raggiungendo la sommità dello strato rigido. Dato che eravamo penetrati entro i sedimenti pliocenici piú antichi, lo strato compatto avrebbe dovuto risalire al Miocene superiore: a circa 6 milioni di anni fa. Nel momento critico, tuttavia, mentre tutti aspettavamo con impazienza di poter osservare il campione successivo, accadde un incidente: il carotiere si incastrò nella colonna di perforazione. Non c'era nulla da fare se non ritirare tutta la colonna di perforazione sul ponte ed estrarre il carotiere dall'estremità delle aste di perforazione. Inoltre ci fu consigliato di trovare un altro punto ove perforare: dove ci trovavamo non potevamo sperare di penetrare a profondità maggiori di quelle già raggiunte.

Né io né Ryan dormimmo quella notte: la passammo ad analizzare i dettagli della perforazione. Il carotiere era finalmente tornato sul ponte e lo trovammo incastrato in un'asta colma di sabbia: posta in un secchio essa fu portata nel nostro laboratorio di bordo. Tutta la mattina del 24 fu impiegata da Ryan e da me per separare dalla sabbia ciottoli delle dimensioni di un pisello; eravamo troppo stanchi per lavorare e troppo eccitati per dormire. Ci occorreva un lavoro semplice per smaltire la tensione. Mentre la mattinata era sul finire, tuttavia, noi fummo sempre piú colpiti da quanto ci capitava di osservare.

I ciottoli sono rari negli oceani. Le frane sottomarine possono generare correnti sottomarine, note col nome di correnti di torbidità, che possono trasportare le sabbie costiere e i ciottoli fin sulle piane abissali. È logico attendersi però che questo tipo di ciottoli sia costituito da tipi differenti di rocce derivate dall'erosione sulla terra emersa. Mentre i ciottoli che stavamo selezionando si accumulavano, ci accorgemmo che essi appartenevano sostanzialmente a tre tipi di rocce diverse: basalto oceanico, fango oceanico consolidato e gesso. Inoltre vi era una fauna nana non comune. Non trovammo quarzo, feldspati, né graniti o rioliti, né gneiss o scisti, non quarziti né arenarie. In sostanza non trovammo nulla che potesse essere identificato come proveniente dalla massa continentale adiacente. I ciottoli non potevano essere stati trasportati dalle correnti di torbidità dalle coste spagnole. Che significato avevano dunque quei ciottoli?

Il giorno 24, verso sera, fu preso in considerazione lo scalpello di perforazione: fra i denti dello scalpello, che si era incastrato nella sommità dell'orizzonte compatto, vi erano sottili frammenti di aggregati anidritici, cioè di solfato anidro di calcio. L'anidrite è un minerale comune nelle evaporiti, che rappresentano i residui inorganici lasciati dall'evaporazione di acque salate; le evaporiti antiche sono comunemente litificate, cioè trasformate in rocce. Cosí avevamo la prima soluzione a uno dei grandi problemi cui si è accennato: l'orizzonte *M*, lo strato compatto che si trova entro il fondo di tutto il Mediterraneo, è un'evaporite del Miocene superiore.

Questa soluzione però non faceva che rendere il mistero ancora piú fitto. Le evaporiti dovrebbero essere depositi di lagune costiere o di deserti. Perché avremmo dovuto trovare una formazione evaporitica al di sotto del fondale del Mediterraneo, a una profondità di migliaia di metri al di sotto dell'attuale livello del mare?

I ciottoli costituirono il filo per giungere alla soluzione del problema. Essi non potevano essere giunti dalla terra emersa ma solo da un oceano disseccato. Era possibile però che il Mediterraneo fosse rimasto isolato dall'Atlantico e si fosse mutato in un deserto nel Miocene superiore?

Ci si può immaginare il graduale ritiro del Mediterraneo e l'aumento progressivo della salinità delle sue acque con la morte di tutti i normali animali marini salvo che per le specie meglio adattabili a condizioni di ipersalinità. Le porzioni di mare interno alla fine si sarebbero mutate in laghi salati, come il Mar Morto, nei quali si sarebbero raggiunte presto le condizioni di saturazione necessarie alla precipitazione dei gessi. Il progredire dell'evaporazione avrebbe alla fine lasciato allo

scoperto anche il fondo del Mediterraneo. Il vulcano sottomarino ove avevamo iniziato a sottoforare sarebbe divenuto un monte vulcanico interamente all'asciutto e i fanghi oceanici deposti sulle sue pendici si sarebbero litificati. Le acque di deflusso da questo fondo oceanico disseccato avrebbero prodotto il non comune deposito di ciottoli che avevamo trovato.

Allora, però, sembrò assurdo avvalare questo racconto sulla base delle prove ancora deboli di cui eravamo in possesso. Le pianure abissali del Mediterraneo superano i 3000 metri di profondità e il bacino ospita piú di 4 milioni di chilometri cubi di acqua. Era impensabile che questo oceano potesse essere scomparso ed essere stato rimpiazzato da una serie di ambienti analoghi a quello del Mar Morto o quello della Death Valley.

Ma era proprio cosí impensabile? Una serie di fatti e di cifre dimostrava che sarebbe stato abbastanza facile rendere completamente asciutto il Mediterraneo. Infatti, il clima di questa regione è arido; la perdita annua d'acqua per evaporazione è di circa 45 000 chilometri cubi. Soltanto un decimo di tale perdita è compensata dall'afflusso di acque meteoriche e di acque dolci dei fiumi. Il Mediterraneo riesce a mantenere la sua salinità normale tramite lo scambio di masse d'acqua con l'Atlantico. Se lo stretto di Gibilterra venisse chiuso oggi stesso, la perdita annuale di acqua per evaporazione non sarebbe compensata, cosicché il Mediterraneo si asciugherebbe nel giro di circa 1000 anni.

Un altro problema: era proprio necessario ricorrere a una spiegazione cosí drammatica della situazione? Molti geologi avevano già pensato che le evaporiti potrebbero essersi depositate anche da acque profonde soprasalate quando il contenuto in sale fosse divenuto maggiore della concentrazione di saturazione. La distribuzione dello strato di evaporiti ci indicava che questa formazione sedimentaria si era deposta nel bacino profondo del Mediterraneo: non era forse possibile che il bacino fosse colmo di acque soprasalate? È ben vero, ma qualunque ipotesi basata su acque profonde soprasalate non potrebbe spiegare la genesi del deposito di ciottoli di cui si è parlato e, in ogni caso, vi sono, come vedremo, prove assai piú convincenti a favore del modello del disseccamento.

Il vocabolo arabo *sebkha* viene usato nelle zone attorno al Golfo arabico per indicare pianure desertiche aride e particolarmente quelle parti di una pianura costiera desertica che si trova leggermente al di sopra del livello di alta marea. Le sebkha divennero oggetto di notevole interesse per i geologi non appena si comprese che alcuni tipi di formazioni rocciose antiche sono praticamente identiche ai sedimenti che si depongono in quell'ambiente; entrambi sono caratterizzati dalla presenza di anidrite nodulare e di dolomite stromatolitica.

Distribuzione delle evaporiti sul fondo del bacino balearico. Essa corrisponde esattamente a quella tipica dei bacini completamente chiusi. I carbonati e i solfati, meno solubili, essendo i primi a precipitare, si trovano nelle zone periferiche del bacino: l'anidrite e il gesso si trovano lungo una stretta fascia anulare appena piú interna alla regione precedente e che delimita la zona piú centrale in cui si deposita il salgemma; quest'ultimo, infatti, è il composto piú solubile e per questa ragione rimane negli ultimi residui acquosi che si raccolgono soprattutto nella parte centrale del bacino, corrispondente alle pianure abissali.

L'anidrite è una varietà di solfato di calcio che si forma a temperatura elevata. Al di sotto dei 35 °C in presenza di una soluzione satura di cloruro di sodio (NaCl) l'anidrite si idraterebbe a formare gesso ($CaSO_4 \cdot 2H_2O$); la temperatura di idratazione sarebbe piú elevata se la soluzione fosse meno salata. Dato che le sorgenti profonde di acque salse raramente superano i 35 °C di temperatura al fondo, l'anidrite si forma oggi quasi esclusivamente come minerale nei sedimenti di sebkha. Dato che questo minerale precipita andando a riempire i minuscoli pori entro i sedimenti che si trovano presso la superficie superiore delle falde acquifere sotterranee, esso tende a crescere sotto forma di noduli irregolari piuttosto che sotto forma di concrezioni quali si possono trovare nei suoli aridi.

Le stromatoliti sono strutture sedimentarie che si rinvengono nelle rocce carbonatiche e che devono la loro genesi alla crescita di alghe (si veda anche l'articolo di N. D. Newell, *La evoluzione delle scogliere*, in «Le Scienze» n. 52, 1972). Le alghe blu-verdi formano infatti un sottile tappeto sulle pianure litorali; dopo ogni tempesta il tappeto viene sepolto sotto una sottile coltre di sedimenti; la crescita delle alghe, però, procede e piano piano si riforma cosí il nuovo tappeto continuo. Questo alternarsi di eventi dà luogo in definitiva allo sviluppo di una roccia con una struttura caratteristica a lamine che viene chiamata stromatolite. Le stromatoliti sono comuni nelle sebkha della penisola arabica nelle quali si sviluppa anche l'anidrite nodulare.

Per tornare al nostro racconto, il 27 agosto ci dirigemmo verso il bacino balearico a sud di Majorca. Il punto ove iniziare la perforazione fu individuato in una zona poco a nord del piano abissale del bacino balearico: il nostro strumento registratore della profondità ci indicò esattamente una profondità di 1417 fathoms, cioè poco meno di 2600 metri. Le aste di perforazione attraversarono circa 300 metri di fanghi non consolidati prima di toccare il tetto dello strato compatto: da questo fu prelevato poi un campione; esso era costituito, come ci si aspettava, da evaporiti del Miocene superiore. Ciò che sorprese di piú, almeno coloro che sostenevano la deposizione profonda di queste evaporiti, fu la presenza di anidrite nodulare e di dolomia stromatolitica (*si vedano le illustrazioni a pagina 23*).

Le stromatoliti non si possono formare in acque profonde poiché il loro sviluppo è condizionato dalla crescita delle alghe che richiede la luce solare. Per di piú, non ci si può attendere che le acque profonde del Mediterraneo si mantengano sempre a una temperatura superiore ai 35 °C, che è necessaria per provocare la precipitazione di anidrite. Le ricerche petrografiche successive condotte da G. M. Friedman, del Rensselaer Polytechnic Institute, confermarono le previsioni: le evaporiti del Mediterraneo si erano depositate su una pianura desertica; tali sedimenti non possono assolutamente essersi depositati entro una massa d'acqua della profondità di alcune migliaia di metri.

Anche metodi di ricerca piú sofisticati condussero alla stessa conclusione. Per esempio, si sa che l'ossigeno presente nei solfati e nei carbonati è costituito da due isotopi: quello piú comune, l'ossigeno-16, e l'isotopo pesante, l'ossigeno-18. Il rapporto tra ossigeno-18 e ossigeno-16 presente in un certo campione può illuminare sulla sua genesi. Le evaporiti precipitate da acqua di mare sono caratterizzate da uno stretto ambito di composizioni isotopiche. Le evaporiti invece depositatesi nelle *playa*, cioè nei laghi del deserto, hanno un piú ampio grado di variabilità. Le analisi dei nostri campioni compiute da R. M. Lloyd dei laboratori di Houston della Shell Oil Company mostrarono un elevato grado di variabilità nella composizione isotopica, fornendo cosí una conferma supplementare delle loro origini in ambiente di playa.

Cosí l'anidrite risultava depostasi su una pianura desertica: ma che pianura desertica doveva essere! Essa è ora sepolta a circa 2700 metri al di sotto del livello del mare. Si dispone oggi però di una prova inequivocabile che il fondo del bacino mediterraneo era cosí profondo?

Si può rispondere affermativamente. La studiosa italiana M. B. Cita della Università di Milano, che partecipava alla spedizione in qualità di paleontologa di bordo, studiò i microfossili presenti nei sedimenti marini al di sopra e al di sotto del livello di anidrite: essa concluse che i sedimenti sono di tipo oceanico normale. Chiaramente il bacino era profondo e colmo di acque marine quando era in co-

municazione con l'Atlantico; esso però si trasformò in una profonda valle quando le comunicazioni con l'oceano si interruppero e il bacino si disseccò. Dato che scoprimmo diversi sedimenti oceanici intercalati con le evaporiti, ne concludemmo che le comunicazioni furono ripetutamente aperte e chiuse durante un intervallo di tempo di circa un milione di anni.

Dopo sei settimane di perforazioni fummo in grado di confermare che lo orizzonte *M* era costituito ovunque da uno strato di evaporiti. Ciononostante, noi avemmo modo di raccogliere campioni di evaporiti soltanto di tipo carbonatico e solfatico e non di salgemma. Qualcuno cominciò a dubitare che nemmeno un pezzo di salgemma sarebbe mai arrivato sul ponte prima di essere completamente sminuzzato e dissolto nel corso delle operazioni di trivellazione. Noi eravamo preoccupati del fatto che avremmo potuto andar cercando nei luoghi sbagliati. La distribuzione dei minerali salini nelle playa desertiche ha un andamento particolare (*si veda l'illustrazione a pagina 24*). I carbonati e i solfati, meno solubili poiché sono i primi a precipitare, si trovano ai margini del bacino, mentre il salgemma, piú solubile, si depone nelle parti piú centrali e profonde del bacino, ove si vanno concentrando le acque via via piú salate.

I nostri primi pozzi erano situati nelle parti periferiche e, seppur debolmente, elevate del bacino; per trovare il salgemma dovevamo cercare al di sotto del piano abissale. Era ormai ottobre e ci stavamo dirigendo verso casa. Un ultimo pozzo fu perforato sul piano abissale a circa 130 chilometri a occidente della Sardegna. Dopo aver attraversato circa 330 metri di fanghi non consolidati, facemmo finalmente il colpo fortunato: dal carotiere apparve un cilindro cristallino, trasparente e scintillante; il suo sapore amaro non lasciava alcun dubbio: avevamo trovato il salgemma a oltre 3300 metri al di sotto del livello del mare.

Intercalati ai livelli di salgemma, vi erano straterelli di silt depositato dal vento. Questo deposito eolico comprendeva quarzo modellato in ambiente subaereo, cosí come frammenti di scheletri di foraminiferi. Queste minuscole creature marine si erano diffuse nel Mediterraneo nel periodo in cui erano prevalse le condizioni marine. Dopo che il Mediterraneo era rimasto isolato e si era disseccato, i loro minuscoli scheletri erano stati sparsi dal vento su tutta la pianura desertica e, in qualche caso, erano caduti entro un lago salato, entrando cosí infine a fare parte della roccia evaporitica che si stava formando. A testimonianza dell'alternarsi di condizioni umide e asciutte della playa, vi sono l'anidrite nodulare e i fanghi disseccati riempiti di evaporiti.

Il salgemma, una volta esaminato al microscopio, mostrò chiara traccia del ripetersi di episodi di soluzione e ricristillizzazione, cosí come accade ora per il sale che si deposita nelle attuali saline costiere della bassa California o in alcune parti della Death Valley. La analogia con quest'ultima può essere sviluppata ulteriormente. In un luogo vicino a quello ove avevamo trovato il salgemma, campionammo dei silt rossi e verdi e ciottoli ben arrotondati tipici delle incisioni fluviali desertiche. Questi erano stati trasportati fino alla base della scarpata continentale, ormai all'asciutto, da improvvise inondazioni sulle montagne della Sardegna ed erano stati depositati sotto forma di conoidi di deiezione al bordo del bacino ove si deponeva il sale. La somiglianza termina quando si rammenta che, mentre la Death Valley si trova all'incirca sul livello del mare, il fondo desertico del Mediterraneo si trovava in una depressione a circa 3300 metri piú in basso.

La nostra perforazione aveva appena scalfito la sommità di un deposito di sale assai spesso: lo strato di evaporiti presente nel Mediterraneo, secondo i dati geofisici, dovrebbe avere uno spessore di 1500-1800 metri. Questa stima è probabilmente non troppo lontana dalla verità, dato che formazioni evaporitiche del Miocene superiore sono presenti anche in Sicilia e raggiungono per l'appunto lo spessore di diverse centinaia di metri. A questo punto, riteniamo che le evaporiti siciliane rappresentino una porzione del fondale del Mediterraneo che è stata sollevata pochi milioni di anni fa da movimenti orogenetici.

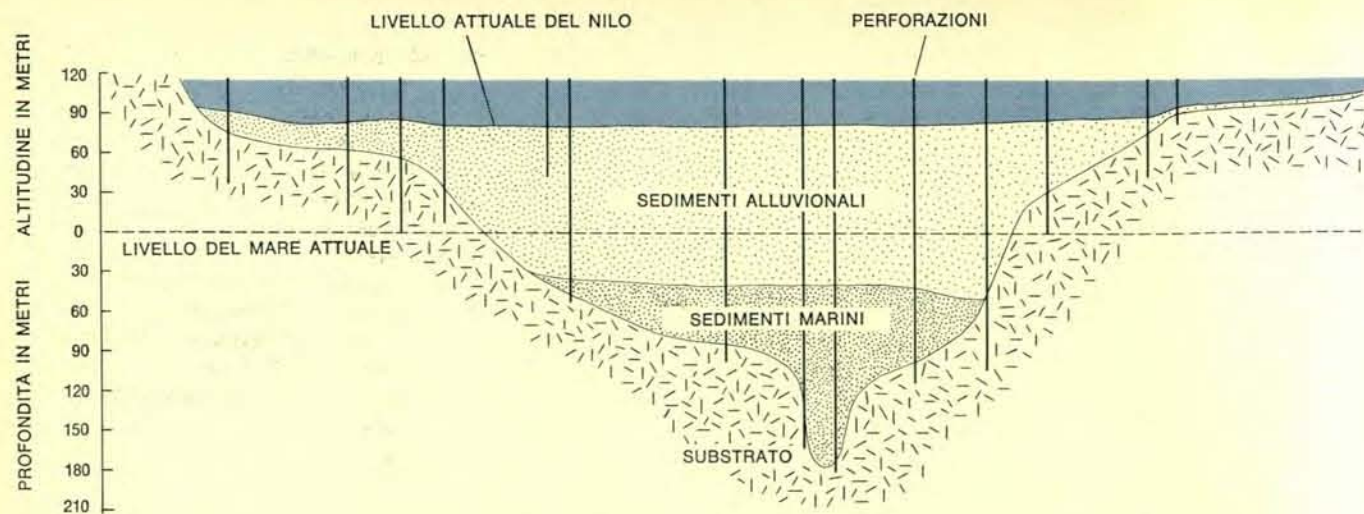
Dopo il sale venne il diluvio. Le prove ricavate dai nostri pozzi sono inequivocabili. Abbiamo infatti ottenuto da tre pozzi diversi, uno nel bacino balearico, uno in quello tirrenico e uno in quello ionico, la testimonianza geologica del fatto che parti diverse del Mediterraneo furono simultaneamente invase e sommerse al di sotto di acque marine profonde verso la fine del Miocene, circa 5,5 milioni di anni fa. Il primo deposito è costituito da una marna grigio scura, dello spessore di 12,5 centimetri, depositatasi quando il bacino stava riempiendosi, seguita da un fango bianco con chiazze locali di fango rosso.

Si può descrivere il Mediterraneo disseccato come una grande vasca da bagno in cui lo stretto di Gibilterra funge da rubinetto. L'acqua di mare si riversò dall'Atlantico attraverso lo stretto, determinando una gigantesca cascata. Se la cascata avesse portato nel Mediterraneo 4000 chilometri cubi di acqua di mare all'anno (cioè piú di 10 volte la portata delle cascate Vittoria), il volume non sarebbe stato sufficiente a rimpiazzare la perdita dovuta alla evaporazione. Per evitare che il mare che andava riempiendosi diventasse troppo salato anche per microfaune resistenti come quelle trovate nelle marne grigio scuro, l'afflusso avrebbe dovuto essere almeno 10 volte maggiore dell'evaporazione. Con un apporto di oltre 40 000 chilometri cubi all'anno, le cascate di Gibilterra sarebbero state 100 volte piú grandi delle cascate Vittoria e mille volte piú grandi di quelle del Niagara. Anche con un afflusso cosí grande sarebbero occorsi piú di 100 anni per riempire tutto il bacino. Che spettacolo impressionante deve essere stato questo per i progenitori africani dell'uomo, ammesso che qualcuno di essi sia stato richiamato dallo scroscio delle acque.

Dal momento in cui si depositarono i fanghi bianchi, il Mediterraneo deve essere stato pieno fino all'orlo. Quindi doveva essere attivo anche l'attuale sistema di scambio con le acque atlantiche. I fanghi bianchi sono un sedimento oceanico tipico, costituito quasi interamente di scheletri microfossili e nannofossili; oltre a questi animali planctonici, se ne sono trovati anche di bentonici. W. E. B. Benson della National Science Foundation, ha esaminato gli ostracodi marini e O. L. Bandy, dell'Università della California meridionale, ha studiato i foraminiferi bentonici. Entrambi hanno concluso che a quel tempo il Mediterraneo era piú freddo e piú profondo di oggi. Le nuove specie bentoniche provenienti dall'Atlantico si erano trascinate sul fondo attraverso il profondo squarcio che si apriva al margine occidentale del Mediterraneo. Alla fine, la fauna bentonica d'acque fredde morí quando lo stretto di Gibilterra fu sufficientemente elevato da impedire l'ingresso delle acque atlantiche profonde; questa è una condizione che si è mantenuta fino a oggi.

Quando abbiamo attraversato con la perforazione la formazione evaporitica, abbiamo incontrato anche un sedimento laminato non comune. Oltre a materiali organici e a minuti cristalli di dolomite, il sedimento conteneva diatomee fossili il cui scheletro è costituito da silice (SiO_2). H. Stradner, dell'Università di Vienna, un altro paleontologo di bordo, riconobbe che alcune di queste diatomee potevano essere vissute soltanto in acque dolci o salmastre. Questo fatto fu confermato piú tardi anche da M. Hajós della Accademia delle Scienze ungherese: essa individuò non soltanto specie planctoniche ma anche specie bentoniche. Piú tardi, in un pozzo perforato nel Mediterraneo orientale a sud di Creta, ci è capitato di raccogliere una fauna a ostracodi del Miocene superiore. Un esperto di ostracodi italiano, A. Decima dell'Università di Palermo, ha stabilito, dopo aver esaminato i campioni che avevamo sottoposto alla sua analisi, che queste minuscole creature possono essere vissute soltanto sul fondo di laghi salmastri.

Questa scoperta ci mise in imbarazzo. Da dove poteva essere venuta tutta quest'acqua dolce? Non vi era alcuna ragione per ritenere che il clima mediterraneo dovesse essere molto piú umido 6 o 7 milioni di anni fa, né si è potuto trovare alcuna prova di un improvviso e drastico cambiamento nel volume di precipitazioni o di evaporazione che avrebbe potuto mutare grandi laghi in altrettante Death Valleys. Il mistero fu risolto quando i colleghi au-

La geografia dell'Europa, 7 milioni di anni fa, era abbastanza differente da quella attuale. La maggior parte dell'Europa nord orientale era coperta da un vastissimo lago costituito da acque dolci o salmastre che si estendeva dalle adiacenze di Vienna verso est fino al mar d'Aral. Questo grande lago interno è stato chiamato, dal geologo francese M. Gignoux, Lac Mer. A quel tempo il Mediterraneo era già separato dall'Atlantico. Poco prima del sollevamento dei Carpazi, avvenuto circa 7 milioni di anni fa, il Lac Mer entrò in comunicazione con il Mediterraneo, fornendo le acque dolci e salmastre necessarie a formare una serie di grandi laghi interni, dei quali restano, come traccia, numerosi depositi di diatomiti. Alla fine i Carpazi si sollevarono e formarono una barriera che chiuse le comunicazioni tra il Lac Mer e il Mediterraneo; quest'ultimo si trasformò in un grande deserto. I grandi laghi si ridussero gradualmente a laghi soprasalati e playas in cui precipitarono i depositi evaporitici.

Profonda gola sepolta scoperta nell'alta valle del Nilo presso Assuan in Egitto. Essa fu scoperta dal gruppo di geologi russi che ha condotto gli studi preliminari e le perforazioni esplorative per la costruzione della grande diga di Assuan. Il punto più profondo della gola si estende fino a oltre 200 metri al di sotto dell'attuale livello del mare; l'incisione è stata con ogni probabilità scavata quando il Mediterraneo era asciutto. Non appena nel bacino del Mediterraneo, all'inizio del Pliocene, tornò nuovamente il mare, anche questa gola fu prima allagata e poi colmata, perlomeno nella sua parte più profonda, di sedimenti marini.

striaco e balcanico ci parlarono del Lac Mer. Sembra infatti che gran parte dell'Europa orientale, durante il Miocene superiore e il Pliocene, sia stata coperta da acque dolci o salmastre. A un certo momento, un enorme lago, chiamato dai francesi Lac Mer, si estendeva da Vienna agli Urali e al mare di Aral; i suoi ultimi discendenti sono il mar Caspio e il Mar Nero. Questo grande accumulo d'acqua, che raccoglieva tutte le precipitazioni della allora umida e fredda Europa nord orientale, era in collegamento con il Mediterraneo durante la prima parte del Miocene superiore, cioè circa 7 o 8 milioni di anni fa. Poco dopo, grandi movimenti della crosta terrestre condussero al sollevamento dei Carpazi e a una radicale riorganizzazione del sistema idrografico. Il Lac Mer trovò allora uno sbocco verso nord e le sue comunicazioni con il Mediterraneo si chiusero. Rimasta isolata dalla sua principale sorgente di acque dolci, l'arida regione mediterranea subí in tal modo il destino del disseccamento.

Dai pozzi perforati nel Mediterraneo orientale sono stati ottenuti diversi campioni del Miocene medio. Con l'aiuto della sezione affiorante in Sicilia, siamo stati in grado di ricostruire la storia del Mediterraneo durante gli ultimi 15 o 20 milioni di anni. Il Mediterraneo era una volta un ampio braccio di mare che metteva in comunicazione gli oceani Indiano e Atlantico. In seguito alla collisione dei continenti africano e asiatico, e alla crescita delle catene montuose del Medio oriente, si interruppero le comunicazioni con l'oceano Indiano. Nello stesso tempo l'Africa stava anche avanzando verso l'Europa e le comunicazioni con l'Atlantico si mantenevano per la presenza di due strette soglie, quella betica nella Spagna meridionale e quella rifea nell'Africa settentrionale. Nei nostri campioni abbiamo avuto modo di osservare il graduale deteriorarsi dell'ambiente mediterraneo, la progressiva stagnazione delle sue acque di fondo, l'inevitabile estinzione degli organismi bentonici, la lotta per la sopravvivenza delle faune planctoniche e nectoniche e l'evoluzione di specie resistenti che potevano sopravvivere in un ambiente caratterizzato da ampie variazioni di salinità. Abbiamo avuto modo di vedere il mutamento da mare interno a una serie di grandi laghi, il loro disseccamento e lo sterminio completo della fauna e della flora al fondo di queste Death Valleys del Miocene, a oltre 3000 metri al di sotto dell'attuale livello del mare. Abbiamo avuto inoltre modo di vedere il diluvio e lo stabilirsi di una nuova dinastia faunistica e il mutamento nella fauna marina dalla quale è andata via via progressivamente individuandosi quella odierna.

La scoperta del deserto del Mediterraneo è stata resa possibile dallo sviluppo delle tecniche di perforazione in mare profondo. Le prime persone che furono informate della scoperta furono gli scienziati a bordo della nave che erano stati prescelti per le loro qualificazioni oltre che per la loro conoscenza della geologia locale. Perciò il peso della scoperta fu interamente apprezzato solo dopo che la *Glomar Challenger* aveva fatto ritorno a Lisbona e i risultati delle perforazioni erano stati comunicati al pubblico tramite conferenze stampa che erano state tenute a Parigi e a New York.

Se il Mediterraneo era stato davvero svuotato, le pianure delle terre circostanti erano state elevati altopiani e le isole alti picchi. La prima conseguenza di un abbassamento del livello dell'acqua sarebbe stato il ringiovanimento del ciclo erosivo dei fiumi e cioè un deciso aumento della loro capacità erosiva. La gola sepolta del Rodano, scoperta 80 anni fa, avrebbe cosí potuto essere stata solo una delle numerose valli che sfociavano nel Mediterraneo. Vien spontaneo chiedersi se per caso lungo le coste del Mediterraneo non vi siano altre valli sepolte.

Poco dopo essere ritornati a terra, Ryan ricevette una lettera dal geologo russo I. S. Chumakov che aveva appreso le nostre scoperte attraverso un articolo del New York Times. Chumakov era uno degli specialisti che l'URSS aveva inviato ad Assuan in Egitto per contribuire alla costruzione della famosa diga. Nello sforzo di trovare la roccia compatta necessaria per l'impianto della diga, furono trivellati 15 pozzi. Con grande stupore degli studiosi sovietici, fu scoperta una stretta e profonda gola al di sotto del corso del Nilo, tagliata per oltre 200 metri al di sotto del livello del mare nel granito (*si veda l'illustrazione qui sopra*). La valle fu scavata circa 5,5 milioni di anni fa e riempita di fanghi pliocenici marini a loro volta ricoperti dalle alluvioni del Nilo. Assuan si trova a circa 1200 chilometri dalle coste del Mediterraneo. Alcuni pozzi scavati nel delta del Nilo, fino a profondità di oltre 300 metri, non riuscirono a raggiungere il fondo dell'antico canyon entro il quale doveva scorrere il fiume. Chumakov valutò che la profondità dell'incisione può qui raggiungere i 1500 metri; egli immagina, in definitiva, un grande e profondo canyon interamente coperto dalle sabbie e dai silt del delta del Nilo.

Chumakov non era stato il solo a essere messo in imbarazzo. I geologi delle compagnie petrolifere operanti in Libia avevano anch'essi avuto la loro parte di sorprese. Primo, i loro sismogrammi avevano registrato delle anomalie: vi erano infatti in profondità nel sottosuolo delle strutture lineari che trasmettevano le onde sismiche a velocità anormalmente elevate. Una perforazione eseguita allo scopo di raggiungere queste anomalie rivelò che esse erano canali sepolti incisi a quasi 400 metri di profondità al di sotto del livello del mare. Vi sono anche indizi geologici che concordano nel raccontare la stessa storia: un'erosione vigorosa da parte dei fiumi e l'improvvisa inondazione da parte delle acque marine all'inizio del Pliocene. F. T. Barr e i suoi collaboratori della Oasis Oil Company che ha sede a Tripoli in Libia, conclusero in un loro rapporto che il Mediterraneo doveva essere stato molte centinaia di metri al di sotto del suo livello attuale quando questi canali erano stati intagliati. Però essi non poterono pubblicare mai il loro manoscritto in una rivista scientifica poiché nessuno se la sentiva di accettare una interpretazione che allora appariva cosí rivoluzionaria.

Valli e canali sepolti di questo tipo sono stati trovati anche in Algeria, Israele, Siria e altri paesi circummediterranei. I fiumi che si gettavano in un Mediterraneo in via di disseccamento non avrebbero inciso solamente la terra ora emersa, ma avrebbero continuato anche attraverso tutta la piattaforma continentale allora esposta e attraverso la scarpata continentale, giú fino al piano abissale che si stava trasformando in una serie di playa. Durante una conferenza tenuta alla Università di Yale, e dedicata alla discussione dei risultati della nostra spedizione, uno studente richiamò la mia attenzione su tale eventualità e mi chiese se in effetti tali canali esistono e sono frequenti nell'area considerata.

La risposta è affermativa. Nel bacino balearico sono stati condotti estesi studi oceanografici da parte degli studiosi francesi. J. Bourcart dell'Università di Parigi annunciò nel 1950 la scoperta di numerosi canyon sottomarini che interessano i margini continentali che si sviluppano al largo delle coste della Francia, Corsica, Sardegna e Africa settentrionale.

I canyon hanno una morfologia tipicamente fluviale e sono riempiti da sedimenti alluvionali. La maggior parte di essi può essere collegata a fiumi tuttora presenti sulla terra emersa e può essere seguita fino alla profondità di 1800-2400 metri al margine del piano abissale. Anche questi sono stati scavati all'inizio del Pliocene. Da allora, canyon di questo tipo sono stati individuati in tutto il Mediterraneo; la loro genesi non fu mai spiegata in modo soddisfacente finché non si arrivò a comprendere che l'intero Mediterraneo era stato, sei milioni di anni fa, un vero e proprio deserto.

Questa idea schiude la porta alla soluzione di altri misteri. Per esempio, si può ora cominciare a capire l'origine delle numerose caverne presenti sulle terre che circondano il Mediterraneo e la peculiare topografia carsica del Carso e della Jugoslavia. Si può anche comprendere il problema lungamente dibattuto delle ragioni per cui la circolazione delle acque sotterranee penetri a 3000 metri di profondità in un'isola oceanica come è Malta. Ma non solo i mutamenti dell'ambiente fisico furono catastrofici, ma anche quelli biologici. G. Ruggieri dell'Università di Palermo, un'autorità nello studio dell'evoluzione della fauna marina di scarsa profondità del Mediterraneo, mi ha inviato di recente la ristampa di un articolo che egli ha scritto nel 1955, nel quale, per spiegare la rivoluzione biologica intervenuta sei milioni di anni fa, aveva ipotizzato il disseccamento del Mediterraneo. Il livello individuato da Lyell, al quale si è già accennato in precedenza, segnò il ritorno delle acque marine in questo bacino interno.

Se si togliessero tutte le acque dal Mediterraneo per riversarle negli altri oceani, il livello medio del mare salirebbe di oltre 10 metri: basterebbe per sommergere la maggior parte delle città costiere. L'ampiezza nel decremento del carico dovuto alla scomparsa delle acque dal bacino del Mediterraneo è stata paragonabile a quella della calotta glaciale fennoscandinava presente in Europa durante le glaciazioni. La risommersione del Mediterraneo deve avere provocato, da un lato, la subsidenza del bacino e, dall'altro, il contemporaneo sollevamento delle terre circostanti.

Geologi e oceanografi cominciano a trovare indizi in questo senso. La presenza di un deserto caldo nel luogo ove ora si trova il Mediterraneo deve avere avuto un notevole effetto climatico. Infatti i paleobotanici europei hanno individuato nell'Europa centrale un mutamento nel clima durante il Miocene superiore, diretto verso condizioni di temperatura e aridità maggiori: le foreste viennesi si mutarono in steppa e le palme crebbero in Svizzera. Il clima tornò umido e freddo con il ritorno delle acque marine nel bacino mediterraneo nel Pliocene; in seguito esso andò via via deteriorandosi: le glaciazioni erano alle porte.

È anche interessante notare che la calotta polare artica cominciò a formarsi nel Miocene superiore. Fu una coincidenza oppure il disseccamento del Mediterraneo ebbe un preciso effetto scatenante sul meccanismo della glaciazione?

La scomparsa di un bacino marino di cosí grandi dimensioni può essere considerato un evento unico nella storia geologica? Probabilmente no. L'esistenza di grandi accumuli evaporitici sembra indicare che devono esserci stati altri episodi di disseccamento di bacini oceanici. Le famose evaporiti dello Zechstein dell'Europa settentrionale possono essere ciò che resta di un bacino interno disseccatosi circa 250 milioni di anni fa. I grandi depositi di sale e di potassio di Alberta e di Saskatchewan, che hanno un'età di circa 350 milioni di anni, possono avere la stessa origine.

Di fatto, la scoperta che un piccolo bacino oceanico può essersi mutato in un deserto, ci ha condotto al riesame di tutto il problema dell'origine dei depositi evaporitici.

I geologi erano continuamente tormentati dalla presenza di grandi depositi salini oceanici sotto il Golfo del Messico, sui fondali dell'oceano Atlantico meridionale al largo delle coste del Congo e dell'Angola e sotto il fondo dell'oceano Atlantico settentrionale al largo delle coste della Nuova Scozia. Si può ora ipotizzare che questi depositi salini si siano formati quando tanto la zona caraibica quanto lo stesso Atlantico erano mari interni che stavano disseccandosi.

Può sembrare ovviamente innaturale immaginare il Mediterraneo come un profondo, asciutto e caldo inferno. Noi stessi eravamo riluttanti finché tutte le altre spiegazioni non hanno fallito: i fatti non lasciano alternative. E, secondo una vecchia massima di Sherlock Holmes, « quando avete escluso l'impossibile, qualunque cosa rimanga, per quanto improbabile, deve essere la verità ».

Malattie e risposte immunitarie

Il meccanismo di difesa del corpo non sempre è utile. In molti casi è lo stesso processo che dovrebbe combattere un virus a provocare il danno connesso con la malattia virale

di Abner Louis Notkins e Hilary Koprowski

In biologia come in fisica è cosa nota che piú a fondo si analizzano dei problemi elementari, piú ci si imbatte in fenomeni paradossali e di difficile interpretazione. Questo è quanto si verifica nello studio del ben noto meccanismo di difesa animale detto risposta immunitaria, per mezzo del quale l'organismo combatte le infezioni e le invasioni da parte di sostanze estranee. Il concetto classico di questo meccanismo è abbastanza semplice: in risposta all'invasione da parte di una sostanza estranea, o antigene, l'organismo produce degli anticorpi specifici che si legano a esso, neutralizzandolo.

Nel caso di attacco diretto da parte di un virus (per esempio nella poliomielite), il virus invade una cellula e utilizza il materiale cellulare per replicarsi. I virus neoformati fanno scoppiare la cellula e vanno a infettare altre cellule. La comparsa tempestiva di anticorpi può impedire il diffondersi dell'infezione e la comparsa di sintomi. In infezioni causate da altri virus, tuttavia, ci sono prove sempre piú numerose che le cellule non sono danneggiate direttamente dal virus in replicazione, ma da una risposta immunitaria specifica che produce i sintomi della malattia. La complessità della risposta immunitaria ai virus è allo studio in numerosi laboratori, compreso il nostro al National Institutes of Health (Notkins) e all'Istituto Wistar di anatomia e biologia (Koprowski). Si sta mettendo insieme gradualmente una interpretazione delle diverse fasi del meccanismo immunitario, e quanto segue è una rassegna del quadro che ne è emerso.

Il fatto che il sistema immunitario possa essere qualche volta responsabile di effetti dannosi è stato suggerito per la prima volta piú di 60 anni fa da Clemens von Pirquet, un pediatra austriaco che fu anche professore alla Johns Hopkins School of Medicine. Von Pirquet osservò che nella « malattia da siero », una malattia che può verificarsi in seguito a iniezione di siero estraneo, il sangue del paziente conteneva proteine estranee e anticorpi contro di esse. Egli formulò l'ipotesi che la combinazione di anticorpi e proteine estranee (antigeni) produceva una sostanza tossica che dava luogo ai sintomi della malattia; orticaria, esantema, dolori alle giunture, insufficienza respiratoria, e nei casi piú gravi, morte. Egli suppose anche che un'interazione di anticorpi con i virus di malattie quali vaiolo e morbillo potesse causare le eruzioni cutanee caratteristiche di queste malattie.

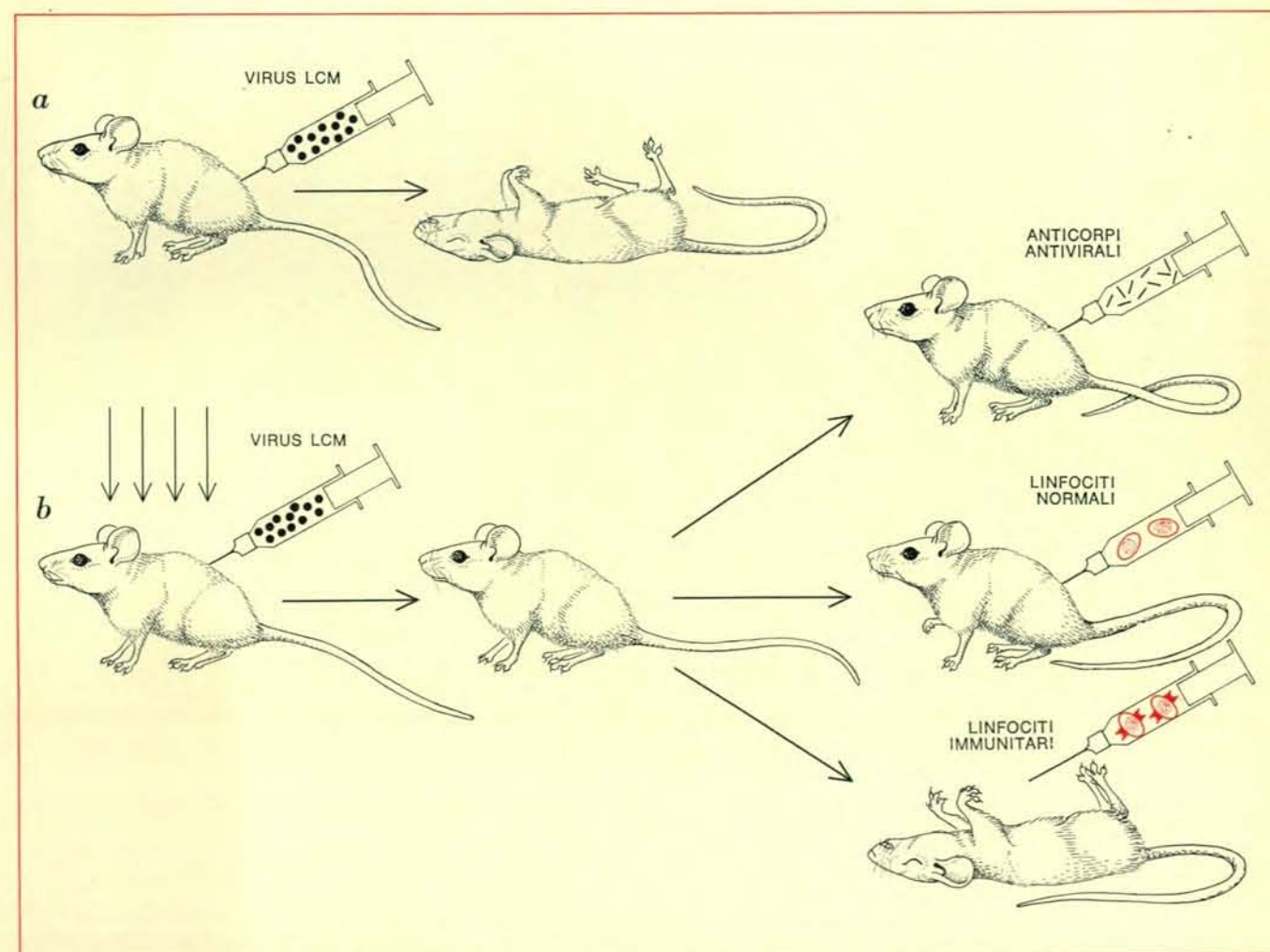
Le ipotesi di von Pirquet secondo le quali la risposta immunitaria a virus potrebbe causare malattie, non ebbero seguito a quel tempo, ma negli anni cinquanta Wallace P. Rowe, un virologo del National Institutes of Health, provò l'ipotesi con una geniale serie di esperimenti. Rowe stava studiando il quadro patologico prodotto da un virus, noto come coriomeningite linfocitaria (LCM) che infetta i roditori, e talvolta l'uomo, e provoca un'infiammazione delle membrane che circondano il cervello (meningite). Egli osservò che sebbene nei topi infettati il virus si moltiplicasse rapidamente in molti organi, gli animali inizialmente non mostravano alcun segno di malattia. Tuttavia, a partire dal sesto giorno da quando i topi avevano cominciato a mostrare una risposta immunitaria verso il virus, si sviluppava la meningite e gli animali morivano. La malattia era provocata dalla loro risposta immunitaria al virus, piuttosto che dal virus stesso? In base all'ipotesi che inibendo la risposta immunitaria si poteva impedire l'insorgere della malattia, Rowe trattò i topi con dosi di raggi X atte a sopprimere la risposta immunitaria. Egli poi infettò sia i topi trattati che i topi di controllo non trattati col virus responsabile dell'LCM. Negli animali irradiati non si sviluppò la meningite, sebbene il virus si moltiplicasse nei loro tessuti con la stessa rapidità riscontrata nei topi di controllo che morivano.

In seguito, un gruppo di ricercatori della Johns Hopkins School of Hygiene and Public Health (Donald H. Gilden, Jerald A. Cole, Andrew A. Monjan e Neal Nathanson) fecero un passo avanti rispetto agli esperimenti di Rowe. Era noto, a quel tempo, che il sistema immunitario risponde alle sostanze estranee in almeno due modi, uno mediato dagli anticorpi e l'altro mediato da uno specifico gruppo di cellule note come linfociti. Questi linfociti sanno riconoscere gli antigeni sulla superficie di cellule estranee e perciò distruggono tessuti come tumori e trapianti di pelle (si veda l'articolo *I marcatori della individualità biologica*, di Ralph A. Reisfeld e Barey D. Kahan in « Le Scienze », n. 49, settembre 1972). Quale di questi fattori era responsabile dell'insorgere della malattia nota come LCM negli esperimenti di Rowe: gli anticorpi o i linfociti? Il gruppo della Johns Hopkins usò dei farmaci per sopprimere la risposta immunitaria nei topi, li infettò con il virus dell'LCM e poi divise gli animali in tre gruppi. Un gruppo riceveva iniezioni di anticorpi anti-LCM, al secondo venivano dati linfociti anti-LCM e al terzo linfociti normali. Gli animali che ricevevano gli anticorpi o i linfociti normali stavano bene, ma in quelli ai quali venivano somministrati i linfociti anti-LCM insorgevano i sintomi della malattia LCM e morivano. Evidentemente nel caso dell'LCM era la combinazione dei linfociti immunitari e del virus che provocava la malattia.

Il gruppo della Johns Hopkins, ampliando le sue osservazioni, trovò che in ratti giovani l'infezione dell'LCM non era fatale, ma danneggiava il cervelletto provocando atassia (incapacità di coordinare i movimenti del corpo). Se, tuttavia, la risposta immunitaria veniva soppressa al momento dell'infezione, negli animali non c'erano sintomi e non c'erano danni cerebellari, anche se il virus continuava a replicarsi nel cervello. Si dimostrò cosí che, come nel caso dei topi, lo sviluppo della malattia nei ratti era mediato immunologicamente. Un'ipotesi interessante che deriva da questi esperimenti è che forse altri disturbi neurologici possono insorgere dalla risposta immunitaria ai virus.

Era quindi tempo di ricercare le ragioni per le quali i linfociti distruggevano le cellule infettate, mentre questo non avveniva per il virus. Alla scopo di studiare questo problema, Duard L. Walker e i suoi collaboratori della Scuola di medicina dell'Università del Wisconsin, passarono da esperimenti su animali a esperimenti su colture di tessuti. Era noto che nell'infettare una cellula alcuni virus inducono la formazione di antigeni virali sulla superficie della cellula. Walker ipotizzò che se questi antigeni venivano riconosciuti da linfociti di animali immunizzati con lo stesso virus, i linfociti potevano attaccare e distruggere le cellule che portano i segni caratteristici dell'infezione. Per verificare questa ipotesi, egli infettò cellule di colture di tessuto con il virus della parotite che induceva la formazione di nuovi antigeni sulla superficie delle cellule, ma non le distruggeva. Quando egli introduceva nelle colture infettate linfociti presi da animali che erano stati immunizzati con quel virus, i linfociti in effetti distruggevano le cellule infettate. D'altra parte, i linfociti di animali che non erano stati immunizzati con il virus della parotite non attaccavano nei nostri esperimenti le cellule infettate.

Altri ricercatori ottennero presto gli stessi risultati in esperimenti con colture di tessuto con il virus dell'LCM e il virus del morbillo. Numerosi gruppi stanno studiando la possibilità che l'interazione tra i linfociti immunitari e gli antigeni virali formati sulla superficie di cellule infettate può spiegare alcuni dei sintomi associati alle malattie virali nell'uomo, compresa l'epatite.

Gary Rosenberg e Paul Farber, nel laboratorio di Notkins al National Institutes of Health, intrapresero un'analisi anche piú dettagliata sul comportamento dei linfociti in risposta agli anticorpi virali. Essi usarono il virus dell'herpes simplex (HSV) che provoca nell'uomo le note lesioni erpetiche. Essi trovarono che quando i linfociti di animali immunizzati con questo virus venivano incubati in una provetta con antigene HSV, l'antigene « dava il via », nei linfociti, alla replicazione del loro DNA e alla divisione. Questa reazione cominciava nel giro di poche ore dopo

Il virus LCM uccide i topi adulti (*a*). Se la risposta immunitaria viene soppressa con radiazioni o farmaci (*b*) il topo vive, ma manifesta un'infezione cronica. Quando la risposta immunologica viene ristabilita iniettando linfociti provenienti da altri animali immunizzati con LCM, i topi muoiono. L'iniezione di anticorpi anti-LCM o di linfociti normali non provoca la morte.

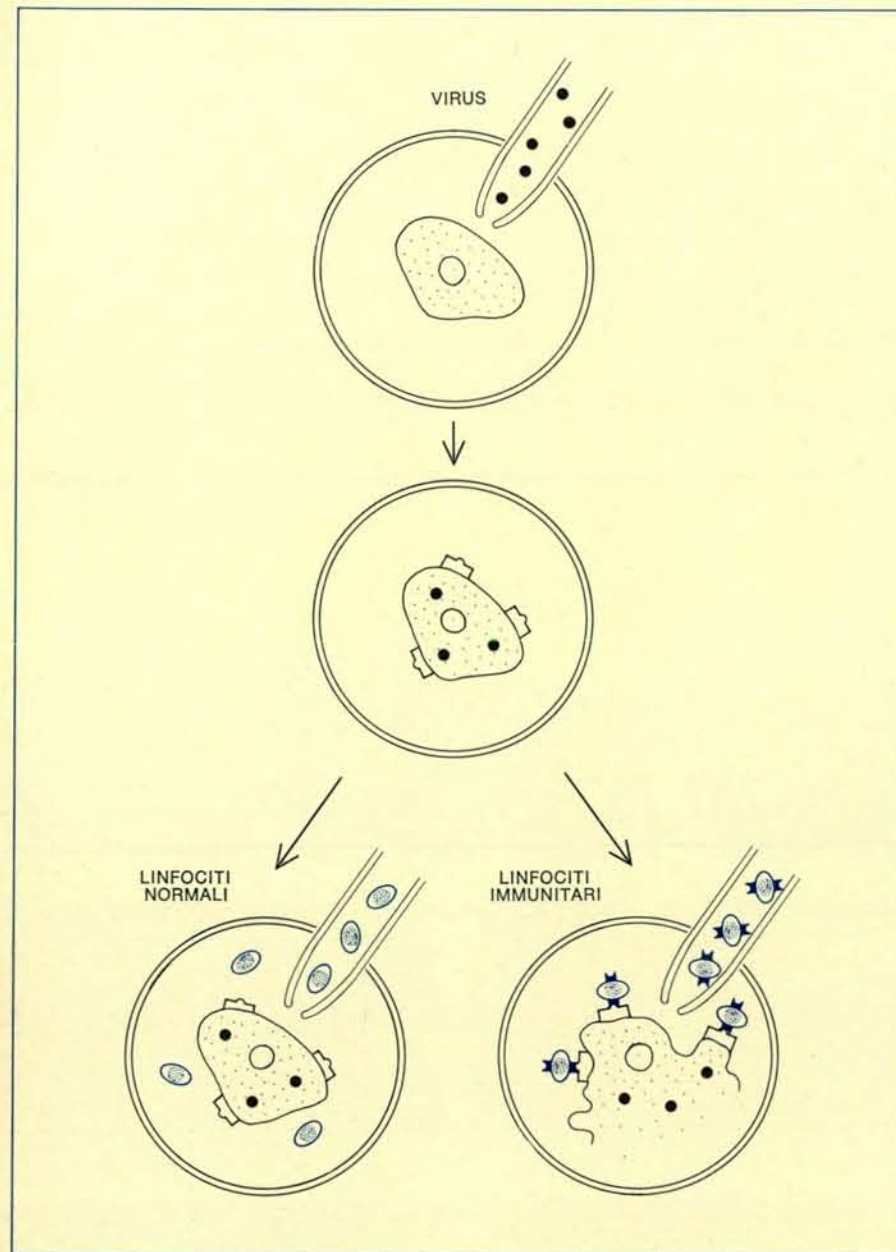
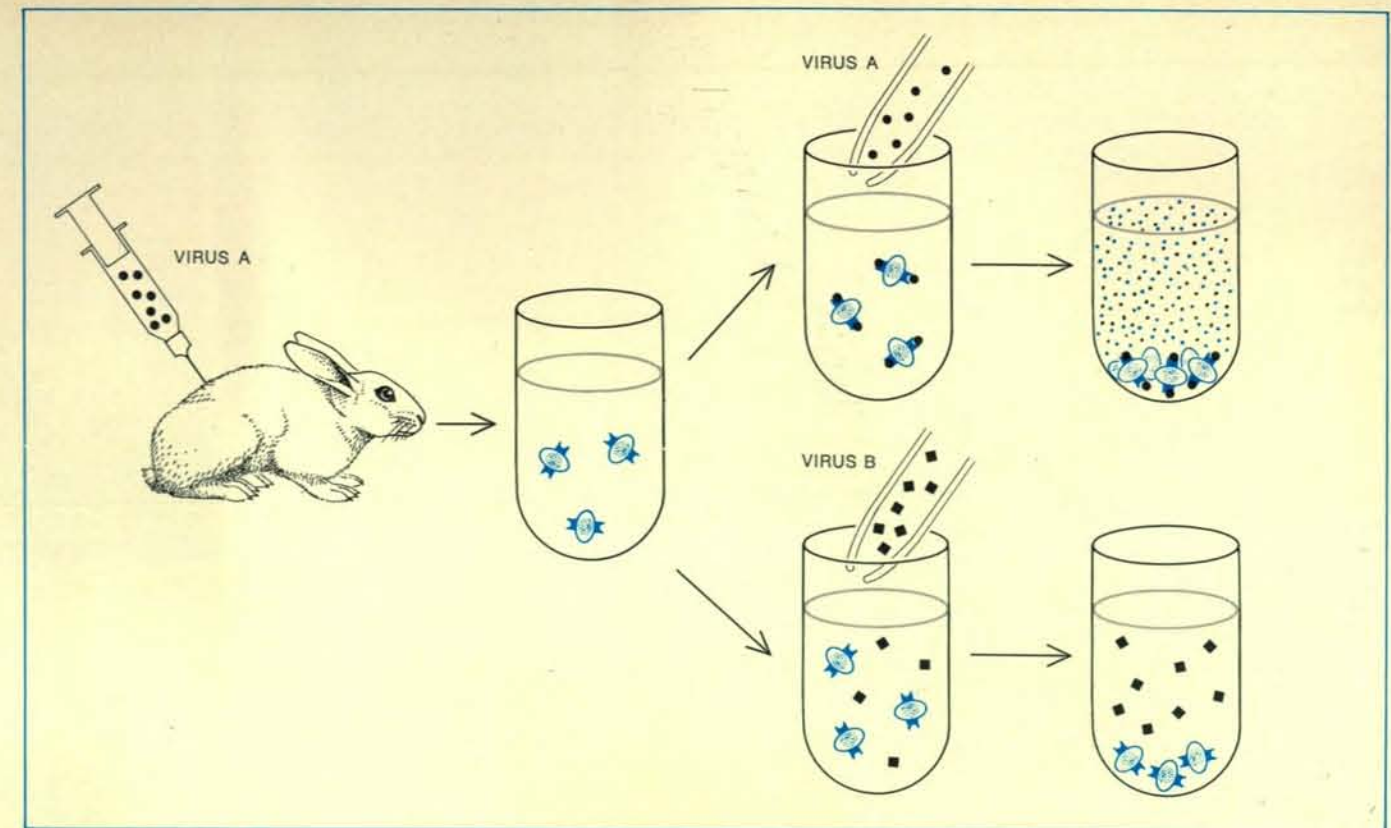
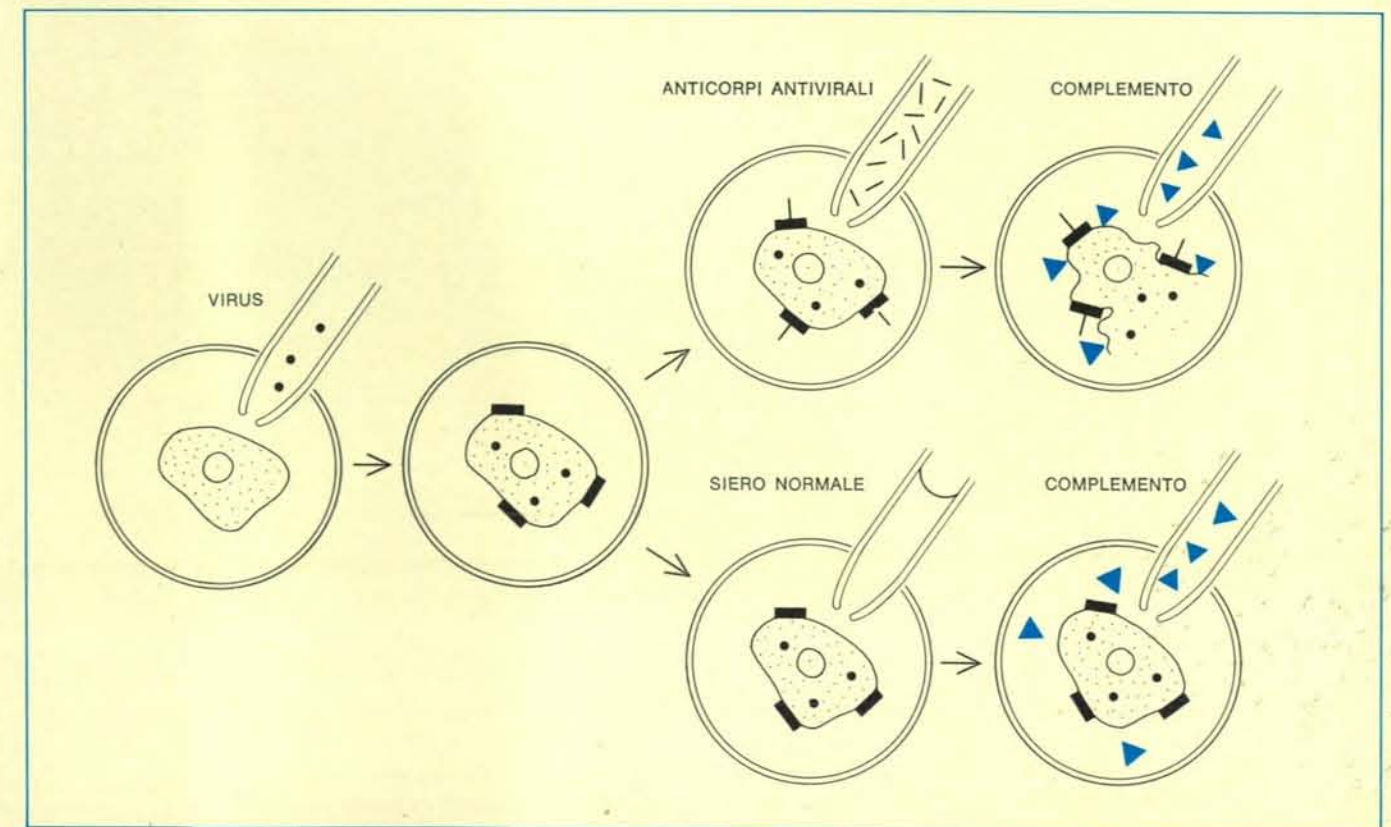
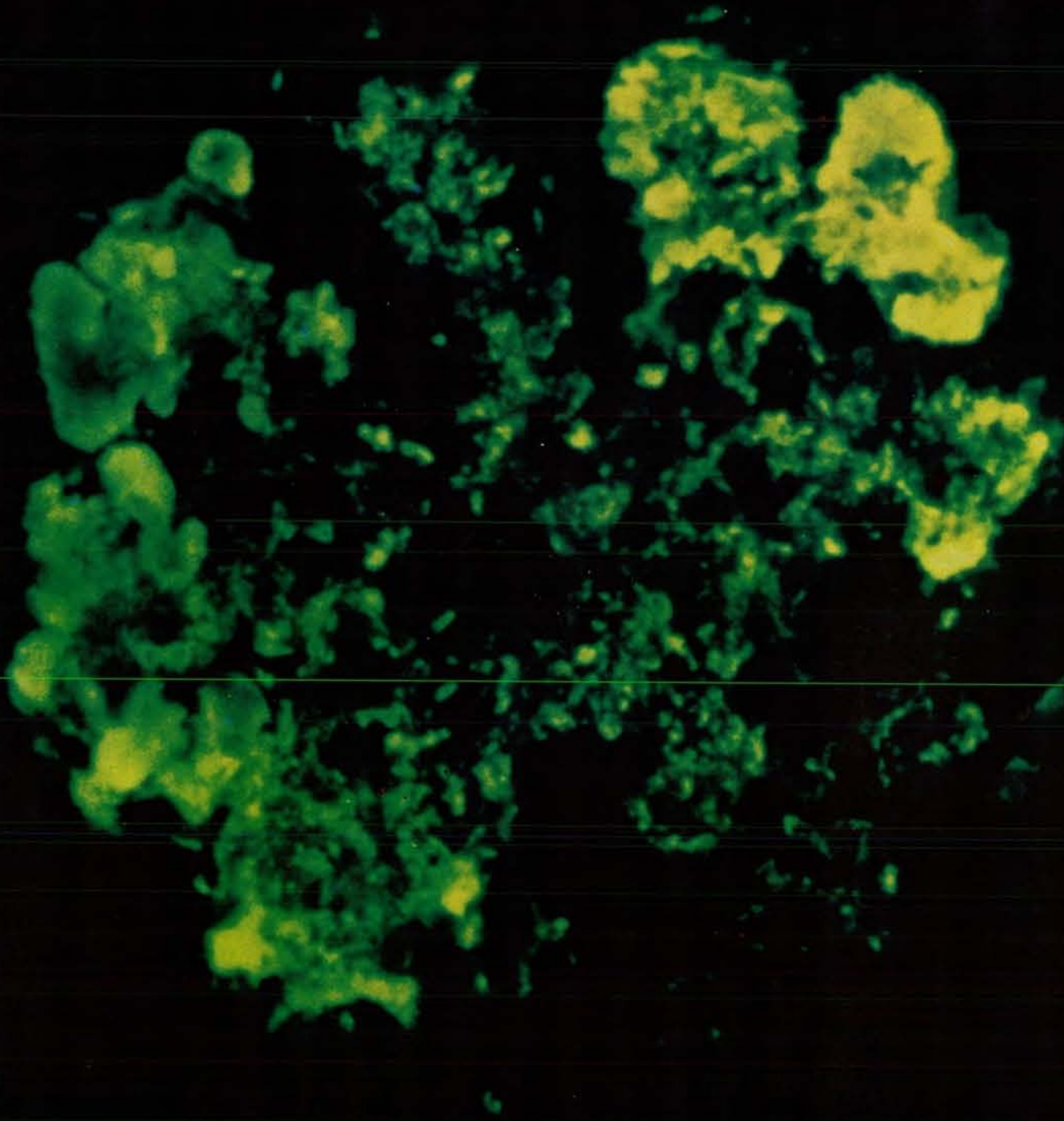
l'esposizione all'antigene ed era piuttosto specifica: i linfociti anti-HSV venivano stimolati soltanto dall'antigene HSV; essi non reagivano affatto agli antigeni di altri virus. Proseguendo gli studi con il virus della rabbia, all'Istituto Wistar, Tadeusz J. Wiktor e Koprowski trovarono che i linfociti di conigli immunizzati con quel virus potevano essere stimolati non solo dal virus completo, ma anche dalle sue subunità.

Ulteriori ricerche mostrarono che quando i linfociti sono stimolati dalla esposizione ad antigeni virali, essi rilasciano potenti messaggeri chimici, o mediatori, che presentano diverse proprietà biologiche e si pensa siano responsabili di alcuni processi infiammatori e alterazioni tissutali associati a molte infezioni virali. Uno di questi mediatori è noto perché attrae i leucociti coinvolti nel processo infiammatorio e un altro perché può trattenere tali cellule nel luogo della infezione. Un terzo mediatore, la linfotossina, è in grado di distruggere cellule infettate e non infettate e un quarto è la ormai ben nota sostanza, interferon, che può inibire la replicazione dei virus. Molto probabilmente, si scoprirà che altri mediatori vengono rilasciati in seguito all'interazione di virus con linfociti; sono stati scoperti in colture di linfociti, mediatori con almeno una dozzina di proprietà biologiche differenti che sono stimolati da antigeni non-virali.

Meccanismo con cui le cellule infettate dal virus vengono distrutte da una risposta immunitaria. Il virus infettante induce la formazione di nuovi antigeni sulla superficie cellulare. L'aggiunta di linfociti normali alla coltura non ha alcun effetto, mentre i linfociti provenienti da un animale immunizzato distruggono le cellule.

Se i linfociti fungono da agenti di distruzione tissutale e di malattie, non possono anche gli anticorpi avere un simile ruolo? Nel laboratorio di Koprowski, Mario Fernandas, Wiktor e Ernest Kuwert cominciarono a studiare la fase anticorpale del fenomeno dell'immunità. Era noto da tempo che l'adesione degli anticorpi agli antigeni sulla superficie di cellule poteva attivare un gruppo di proteine del siero, noto come complemento, per distruggere le cellule. Era anche noto che in alcune circostanze il virus della rabbia poteva indurre nuovi antigeni sulla superficie di cellule in colture di tessuti senza distruggerle. I ricercatori dell'Istituto Wistar aggiunsero alle colture infettate anticorpi antirabbia o complemento, o entrambi. Né gli anticorpi né il complemento da soli erano nocivi alle cellule, ma quando erano aggiunti insieme, la combinazione distruggeva le cellule infettate dalla rabbia.

Recenti esperimenti suggeriscono che gli anticorpi e il complemento possono anche contribuire alla distruzione di cellule infettate con alcuni virus che notoriamente distruggono i tessuti. Charles Wohlenberg, Arnold Brier e Joel Rosenthal del National Institutes of Health dimostrarono che gli antigeni virali sulla superficie di cellule infettate con influenza, morbillo, vaiolo bovino e HSV rendevano le cellule vulnerabili alla distruzione da parte di anticorpi specifici e complemento molto prima che queste cellule venissero distrutte come diretto risultato della moltiplicazione del virus. È molto probabile, perciò, che nel corpo i sintomi e gli altri effetti di queste malattie siano prodotti da una cooperazione tra i processi immunologici e il virus.

Come avviene che l'interazione del complemento con l'anticorpo produce danni e infiammazioni tissutali nell'animale infetto? Ralph Snyderman del National Institutes of Health eseguì un esperimento che suggeriva una risposta verosimile. Egli trovò che quando complemento e anticorpi verso l'HSV venivano aggiunti a una coltura contenente cellule infettate con quel virus, veniva liberato un mediatore da uno dei componenti del complemento. Questo mediatore fu identificato come quello che, come precedentemente dimostrato, aumenta la permeabilità dei vasi sanguigni e attrae i globuli bianchi. Sebbene sia noto che i globuli bianchi sono importanti nella difesa con-

La specificità dell'effetto dei linfociti immunitari è stata dimostrata immunizzando conigli con un virus e quindi mettendo in contatto i loro linfociti con lo stesso virus e con un altro virus. I linfociti nel primo caso venivano stimolati a sintetizzare nuovo DNA e a dividersi; spesso inoltre producevano mediatori, alcuni dei quali distruggevano i tessuti normali.

La risposta immunitaria mediata da anticorpi viene dimostrata in una coltura di tessuto. Nella prima fase gli anticorpi antivirali riconoscono e si legano agli antigeni, indotti dal virus, sulla superficie delle cellule infettate. Il complemento reagisce con i complessi antigene-anticorpo e le cellule vengono distrutte. Né gli anticorpi né il complemento da soli distruggono le cellule.

tro certe infezioni si sa anche che essi contengono enzimi che causano la distruzione dei tessuti. Spesso l'iniezione di anticorpi specifici in animali infettati con il virus, ha l'effetto di aumentare il numero di globuli bianchi e l'entità del danno tissutale negli organi infettati. Sembra quindi che la sindrome da shock, a volte fatale, che accompagna la « dengue », una malattia virale dell'Asia sudorientale, possa essere mediata da anticorpi e complemento. Scott B. Halstead della Scuola di medicina dell'Università delle Hawaii e Philip K. Russel del Walter Reed Army Institute of Research, che proposero originariamente l'idea, pensano che i mediatori liberati dal complemento possano essere uno dei fattori responsabili dell'aumento della permeabilità dei vasi sanguigni e del conseguente shock che distingue questa malattia. Molti ricercatori stanno tentando di raccogliere prove a favore di questa ipotesi.

MECCANISMI	PATOLOGIA
1 ANTIGENI INDOTTI DA VIRUS SULLA SUPERFICIE CELLULARE a INTERAZIONE CON LINFOCITI IMMUNITARI b INTERAZIONE TRA ANTICORPI ANTIVIRALI E COMPLEMENTO	DISTRUZIONE DI CELLULE INFETTATE
2 ATTIVAZIONE DI MEDIATORI DA PARTE DI LINFOCITI IMMUNIZZATI E DI COMPONENTI DEL COMPLEMENTO	INFIAMMAZIONI, REAZIONI ALLERGICHE, DISTRUZIONE DI CELLULE
3 FORMAZIONE DI COMPLESSI CIRCOLANTI VIRUS-ANTICORPO	MALATTIA DA COMPLESSO IMMUNITARIO
4 RISPOSTA IMMUNITARIA VERSO ANTIGENI DELLA CELLULA OSPITE ALTERATA O DEPRESSA A OPERA DEL VIRUS	REAZIONI AUTOIMMUNITARIE
5 INFEZIONE DI CELLULE DEL SISTEMA IMMUNITARIO	INIBIZIONE O AUMENTO DELLA FUNZIONE IMMUNITARIA

Sono qui riassunti i meccanismi di diverse alterazioni patologiche dovute a risposte immunitarie che sono provocate da una infezione virale e vengono trattate nel testo.

Fin qui abbiamo discusso principalmente i meccanismi mediante i quali la risposta immunitaria ad antigeni virali sulla superficie di cellule infettate può causare danni tissutali. Gli studi di von Pirquet che risalgono ai primi del 1900, sulla malattia da siero, suggerivano un meccanismo differente, cioè quello che comporta la combinazione dell'antigene e dell'anticorpo. Si dovette giungere agli anni 50 perché si ottenessero prove sicure. Frederick G. Germuth jr. e i suoi collaboratori all'Ospedale John Hopkins e Frank J. Dixon della Scuola di medicina dell'Università di Pittsburgh trovarono che si poteva produrre la malattia da siero nei conigli iniettando combinazioni, preparate in provetta, di proteine estranee e degli anticorpi relativi. Accertarono anche che alcuni di questi complessi iniettati, in circolo nel sangue di coniglio, venivano bloccati nei capillari renali e producevano infiammazione, perdita delle funzioni renali e sintomi caratteristici di quella malattia nota nell'uomo come glomerulonefrite.

Alla luce delle prove che la malattia renale era causata da un complesso antigene-anticorpo, essa fu chiamata malattia del « complesso immunitario ». Sebbene la sindrome immunitaria potesse essere facilmente prodotta negli animali in laboratorio, il fatto che essa potesse insorgere anche in risposta a infezioni virali in condizioni naturali non fu dimostrato che una decina d'anni dopo. Non fu facile identificare i complessi virus-anticorpi, e fu piú facile affrontare il problema quando fu elaborata una nuova tecnica a questo scopo.

Per molti anni al National Institutes of Health si era studiato un virus insolito chiamato virus della lattico-deidrogenasi (LDV) perché esso aumenta il livello di quell'enzima nel sangue. La inoculazione del virus nei topi produceva grandi quantità di virus infettante nel sangue e un'infezione cronica senza alcuna indicazione di risposta immunitaria. Si suppose che gli animali fossero incapaci di formare anticorpi contro il virus, una situazione nota come tolleranza immunologica. Tuttavia, nel laboratorio si individuò una tecnica molto sensibile per mettere in evidenza le combinazioni virus-anticorpo, e con questo tipo di tecnica (*si veda l'illustrazione in alto a pagina 38*), si scoprí che i topi infettati formavano anticorpi per l'LDV. La ragione per cui questo fatto non era stato individuato prima, era che sebbene l'anticorpo si fosse combinato con il virus, il complesso risultante rimaneva infettante e quindi non poteva essere distinto dal virus stesso.

Altri ricercatori continuarono a dimostrare che i complessi infettanti, virus-anticorpo, erano infatti caratteristici di diverse infezioni virali croniche. Piú importante è il fatto che presto divenne evidente che alcune di queste infezioni croniche sfociavano in glomerulonefrite. John E. Hotchin del Dipartimento di sanità dello stato di New York ad Albany, aveva osservato molti anni prima che l'infezione prodotta dal virus dell'LCM non provocava la morte dei topi neonati, come accadeva invece negli adulti; si instaurava invece un'infezione cronica che si concludeva con glomerulonefrite. Come il virus dell'LCM produce disturbi renali non fu chiarito fin quando Michael Oldstone e Dixon, che lavoravano allora alla Scripps Clinic and Research Foundation a La Jolla in California, mostrarono che il virus dell'LCM esisteva nel sangue di animali cronicamente infetti con un complesso infettante virus-anticorpo e che i reni contenevano grandi quantità di antigeni LCM, anticorpi anti-LCM e complemento.

L'esatta sequenza degli eventi che si verificarono all'insorgere della malattia immunitaria dovuta a complessi virus-anticorpo, è ancora solo parzial-

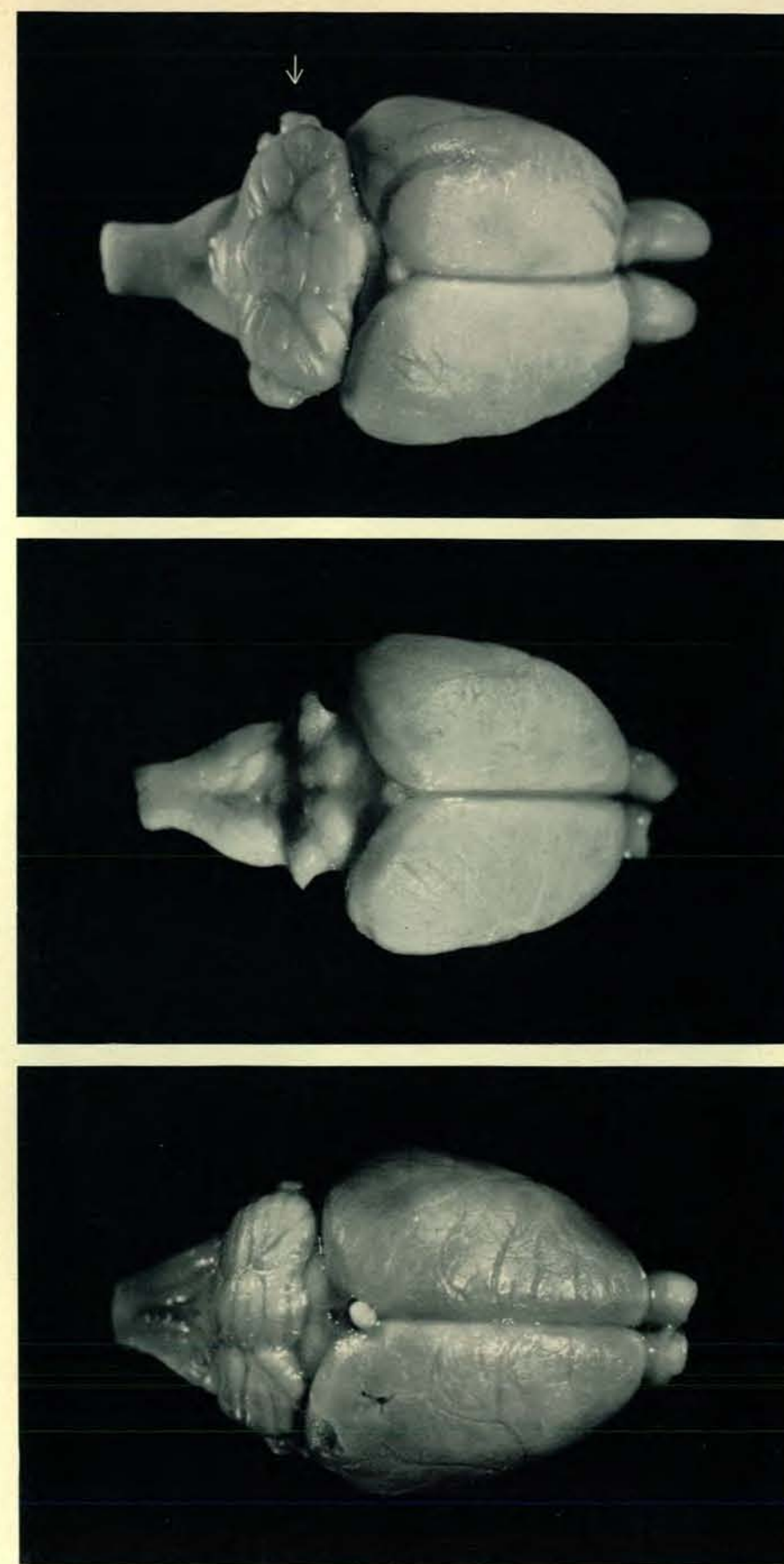
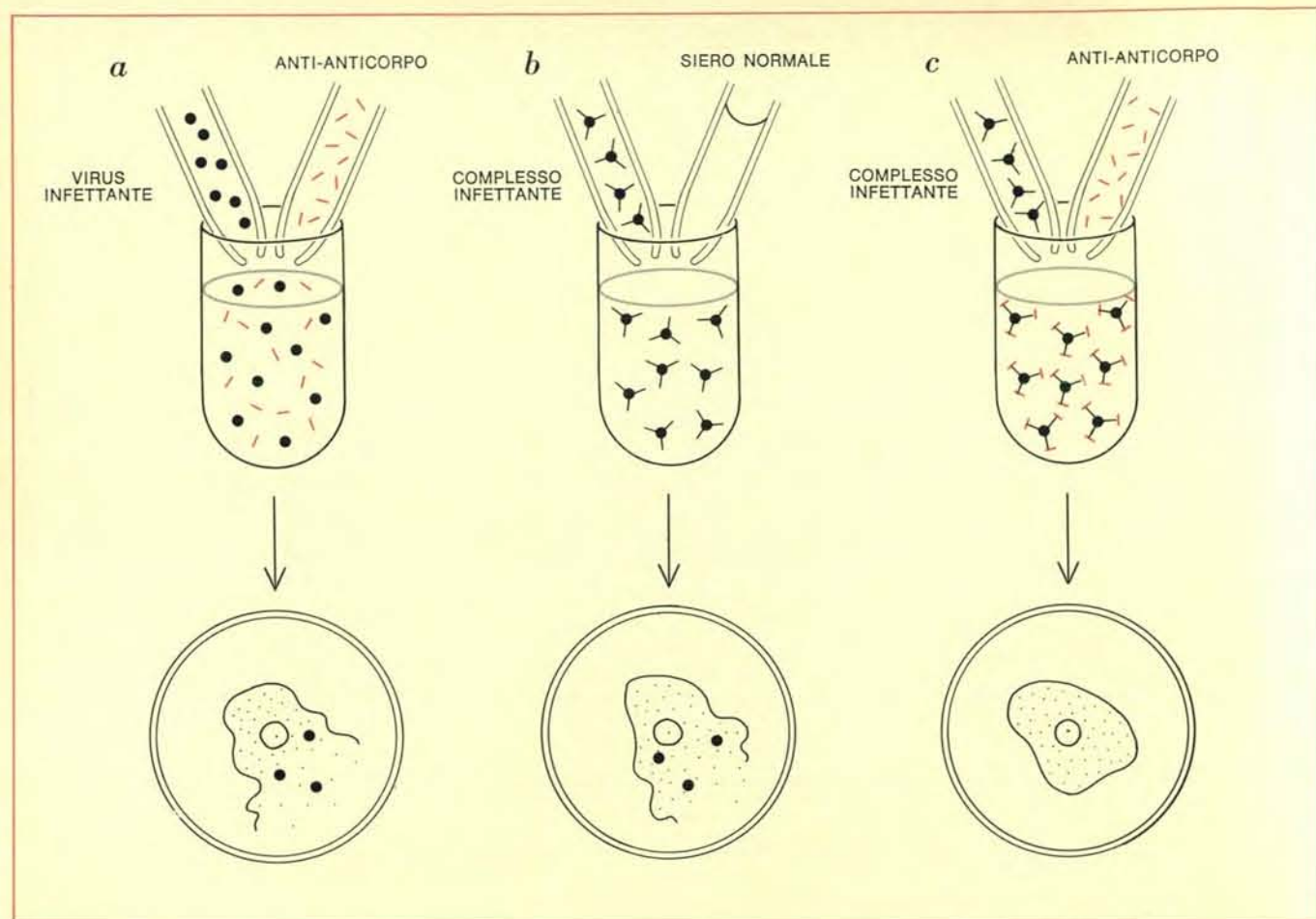
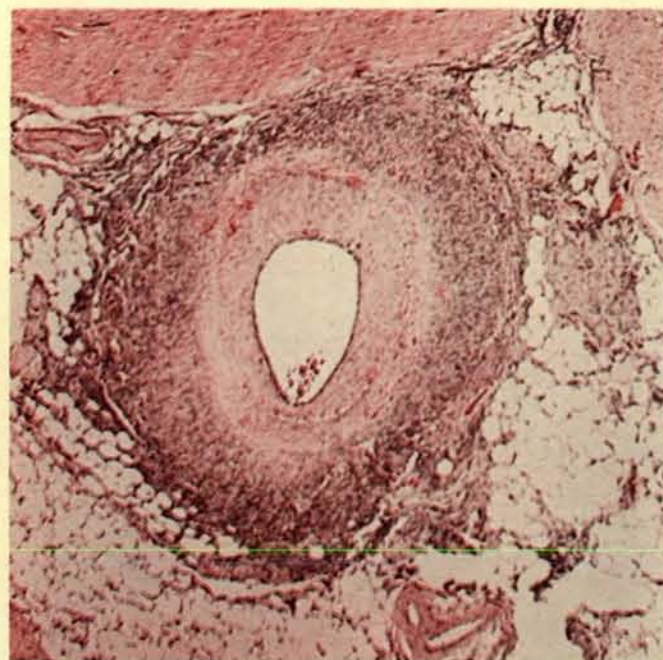
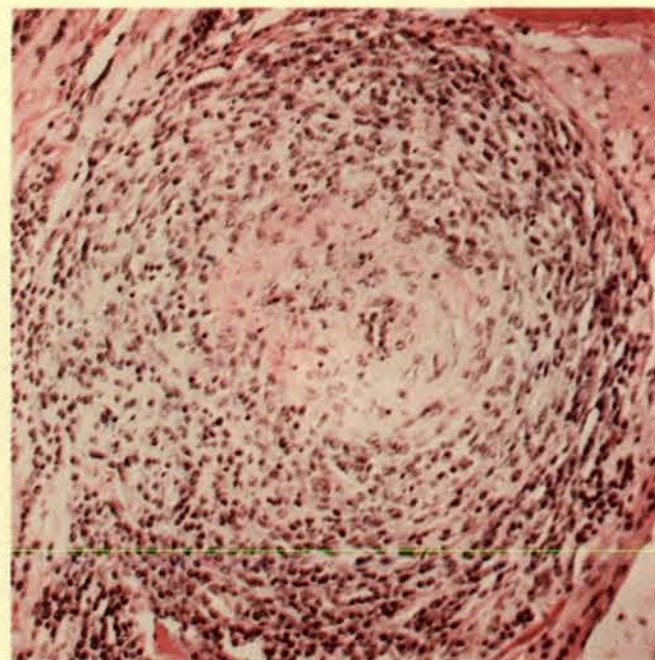
I complessi immunitari si formano quando un anticorpo antivirale si combina con un virus e lega il complemento. I complessi vengono rilevati mediante una tecnica in cui un anticorpo per il complemento viene marcato con una sostanza fluorescente e incubato col tessuto; se sono presenti dei complessi, l'anticorpo si lega a essi e appare fluorescente in luce ultravioletta. La microfotografia della pagina a fronte dimostra la presenza di complessi nei reni di visoni infettati con virus Aleutino e affetti da glomerulonefrite.

L'infezione di topi con virus leucemico diminuiva la loro capacità di produrre anticorpi. I topi, alcuni dei quali erano stati infettati col virus, erano immunizzati con globuli rossi di pecora (*a*). Cellule della milza dei topi venivano stratificate su globuli rossi di pecora (*b*) e veniva aggiunto il complemento (*c*). Solo le cellule di milza nelle quali era stata indotta la formazione di anticorpi verso i globuli rossi di pecora erano in grado, in presenza di complemento, di distruggere i globuli rossi circostanti. Il conteggio dei globuli rossi distrutti indica che la milza di topi infettati conteneva cellule di anticorpi in misura minore rispetto a quanto avveniva nei topi sani.

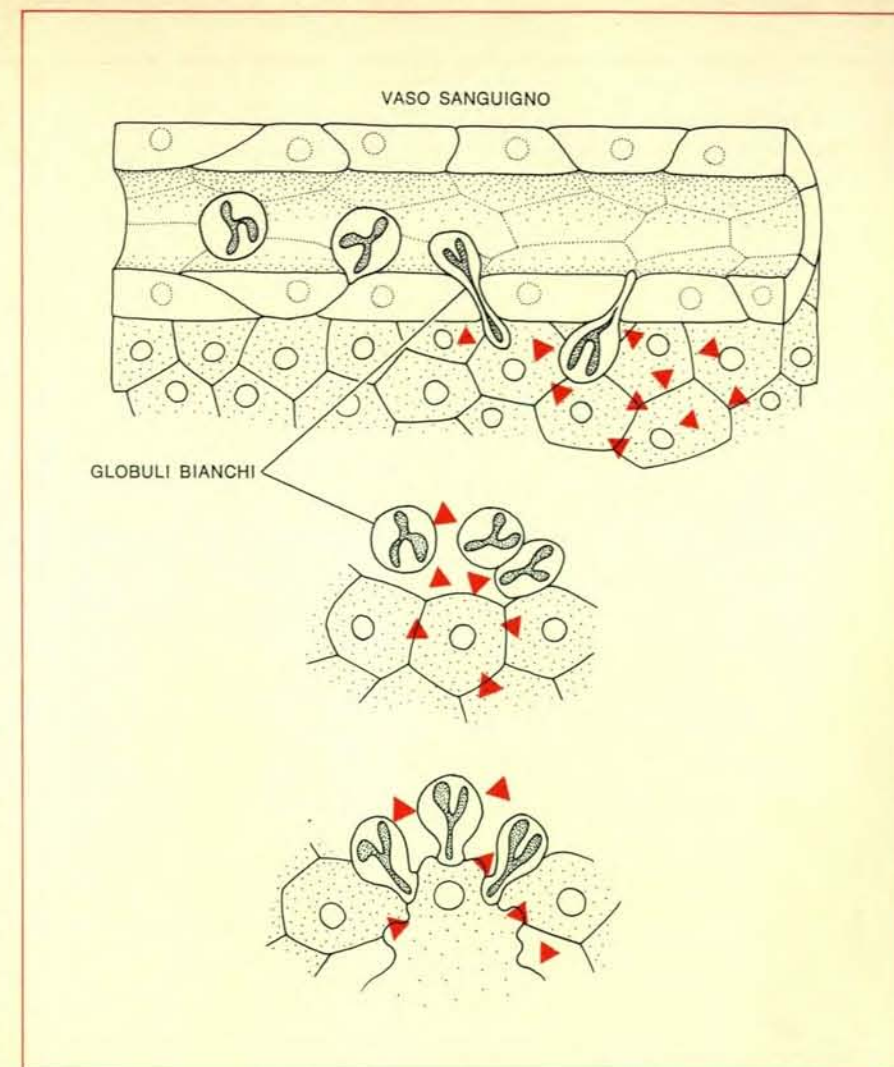
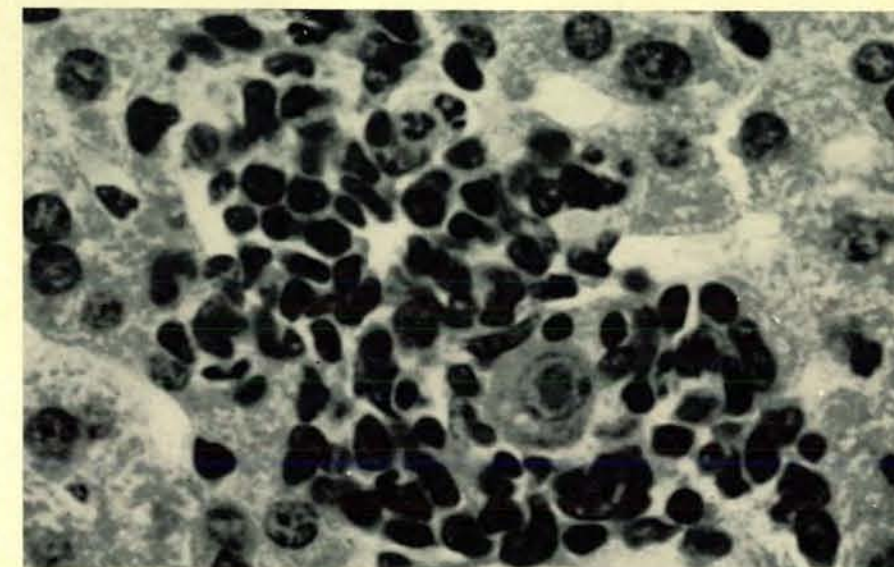
mente chiarita. Probabilmente quando i complessi vengono bloccati nei reni o sulle pareti dei vasi sanguigni, essi attivano i componenti del complemento; si generano mediatori, ha luogo l'infiammazione e gli enzimi che provocano alterazioni tissutali vengono rilasciati dai globuli bianchi. Tuttavia ci sono ancora molti quesiti senza risposta. Perché il virus dell'LCM dà luogo a una grave glomerulonefrite mentre l'LDV dà luogo solo a una forma debole? Perché alcuni animali sviluppano una glomerulonefrite quando sono esposti a un certo virus mentre per altri animali questo non avviene? Esistono fattori genetici che regolano la suscettibilità al virus o la risposta immunologica verso di esso?

In ogni caso, ogni anno, migliaia di persone vengono colpite da glomerulonefrite e quindi l'aver scoperto che i complessi virus-anticorpo producono la malattia negli animali ha intensificato l'interesse per lo studio di tali complessi e il loro eventuale ruolo nella malattia dell'uomo. Recentemente si sono cominciate ad accumulare prove secondo le quali nell'uomo il virus dell'epatite circola nel sangue come un complesso immunitario e alcune delle manifestazioni di questa malattia, compresa la alta incidenza di artrite che l'accompagna, possono essere dovute a questo complesso. Se si vedrà che la risposta immunitaria ai virus è in effetti responsabile della glomerulonefrite o di altre malattie da complesso immunitario nell'uomo, allora per la terapia sarà essenziale controllare gli effetti negativi della risposta immunitaria. Negli animali, molte delle manifestazioni di queste infezioni croniche, comprese la glomerulonefrite, possono essere evitate o ridotte mediante l'uso di farmaci che sopprimono la risposta immunitaria.

In numerosi laboratori è allo studio la possibilità che virus e risposta immunitaria possano essere implicati anche in malattie autoimmunitarie, come l'artrite reumatoide e il lupus eritematoso. Una malattia autoimmunitaria è quella nella quale il corpo tratta il proprio tessuto come un antigene estraneo e produce anticorpi che attaccano il tessuto. Per quale motivo l'organismo improvvisamente si rivolge contro i propri tessuti? Si sospetta da lungo tempo che le infezioni virali siano un fattore scatenante. Sono state suggerite numerose ipotesi per spiegare come un virus possa volgere il sistema immunitario contro le cellule stesse dell'ospite. L'infezione virale può smascherare o liberare un antigene potenziale che si annida nelle cellule isolato dal sistema immunitario. Oppure un antigene virale può combinarsi con una proteina che si trova abitualmente sulla superficie della cellula e cosí formare una nuova sostanza « estranea ». Un'altra possibilità è che un'infezione virale possa attivare dei geni nella cellula la cui formazione è solitamente repressa, facendo sí che essi comincino a produrre nuove sostanze che agiscono come antigeni. Wanda Baranska e Wojciech Sawicki, all'Istituto Wistar, hanno ottenuto dati a favore di quest'idea in esperimenti con uova ed embrioni di topo. Essi dimostrarono che un antigene che era presente nel piú precoce stadio embrionale dell'animale non poteva essere rilevato negli adulti, ma che quando le cellule adulte venivano trasformate in cellule tumorali a opera del virus noto come SV-40, l'antigene embrionale riappariva sulla superficie della cellula. Apparentemente in queste cellule di topo trasformate i geni erano di nuovo in grado di regolare la formazione dell'« antigene embrionale ».

Ancora un'altra possibilità è che una infezione virale faccia in modo che le cellule del sistema immunitario si comportino in modo anormale e quindi producano anticorpi contro qualche tessuto dell'ospite. Sebbene manchino prove a favore di questa ipotesi, ne esistono alcune, derivanti in particolare da esperimenti su animali, secondo le quali certe infezioni virali reprimono la funzione delle cellule del sistema immunitario. E ancora, fu von Pirquet che per primo osservò che la reattività al test della tubercolina (una risposta immunitaria) diminuiva in pazienti infettati dal virus del morbillo. Tuttavia, il problema dell'effetto dei virus sulla funzione immunitaria non ha ricevuto molta attenzione fino al 1936, quando Robert A. Good e i suoi colleghi al College of Medical Sciences dell'Università del Minnesota, mostrarono che in topi infettati con un virus che provoca leucemia veniva notevolmente diminuita la capacità di formare anticorpi contro sostanze estranee. Altri studi, in modo particolare quelli di Walter Ceglowski e Herman Friedman dell'Albert Einstein Medical Center di Filadelfia dimostrarono che non solo la quantità di anticorpi nel sangue era diminuita, ma anche il numero di cellule capaci di formare anticorpi era ridotto fino al 99 per cento (*si veda l'illustrazione nella pagina a fronte*). Inoltre, la risposta immunitaria era diminuita entro pochi giorni dall'infezione, molto prima che negli animali si sviluppassero segni visibili di leucemia.

Cervelli di ratto che mostrano gli effetti della risposta immunitaria all'infezione LCM dimostrata da Andrew A. Monjan, Gerald A. Cole e Neal Nathanson della Johns Hopkins School of Hygiene and Public Health. Il cervelletto si trova alla base del cervello (*in alto*). L'infezione LCM provocava gravi lesioni cerebellari (*al centro*), alterando la capacità di coordinare i movimenti. Invece, in tutti quei casi in cui la risposta immunitaria cellulare veniva soppressa non venivano osservate lesioni cerebrali (*in basso*).

I complessi infettanti virus-anticorpo possono essere rilevati con una tecnica che impiega un anticorpo che riconosce un anticorpo antivirale come antigene: un « anti-anticorpo ». Quando l'anti-anticorpo viene incubato col virus non trattato, il virus mantiene la sua capacità infettante (*a*). Quando il virus con l'anticorpo antivirale attaccato alla sua superficie (complesso infettante virus-anticorpo) viene trattato con siero normale, il complesso rimane infettante (*b*). Quando, però, questo complesso viene trattato con l'anti-anticorpo, quest'ultimo si attacca all'anticorpo del complesso neutralizzando il complesso stesso (*c*).

In alcune malattie virali da complesso immunitario si verifica una arterite. Le microfotografie dell'arteria coronaria di visone infettato da virus Aleutino, mostrano una leggera arterite (*a sinistra*) e a maggiore ingrandimento, una grave arterite con infiltrazione di leucociti e ostruzione del vaso (*a destra*). Le arterie lese contenevano antigeni virali, anticorpi e complemento.

Fu presto chiaro che anche i virus che non producono leucemia potevano alterare la funzione immunitaria. Richard J. Howard e Stephan E. Mergenhagen del National Institutes of Health dimostrarono questo fatto nei test di reattività a innesti di pelle estranea in topi infetti con LDV. Questi animali rigettavano i trapianti in percentuale inferiore rispetto agli animali non infettati. Inoltre ci sono prove che il rigetto di tumori trapiantati è rallentato anche da infezioni virali del sistema immunitario.

Un curioso intreccio nella già complicata storia dell'immunità e dei virus è la scoperta che i linfociti che costituiscono la piú importante difesa del corpo contro i tumori possono fungere in realtà come agenti per l'induzione di tumori. È noto da tempo che un'incidenza insolitamente alta di linfomi (tumori delle ghiandole linfatiche) si verifica in animali o pazienti che subiscono una stimolazione cronica di queste ghiandole come risultato di disordini autoimmunitari o rigetto di un trapianto come quello dei reni. È anche noto che certi virus, compreso il virus della leucemia dei topi e il virus della mononucleosi infettiva dell'uomo si trovano allo stato latente nelle cellule linfoidi. Sulla base di queste osservazioni, gruppi guidati da Martin S. Hirsch e Paul H. Black dell'Harvard Medical School e Robert S. Schwartz della Scuola di medicina della Tufts University hanno condotto degli esperimenti per vedere se la stimolazione degli elementi linfoidi del sistema immunitario può risvegliare il virus latente. Allo scopo di stimolare il sistema immunitario essi esponevano i topi a trapianti estranei, e trovarono che questo serviva ad attivare il virus della leucemia che era stato prima latente nelle cellule linfoidi. Le scoperte suggeriscono che una tale attivazione del virus della leucemia a opera di linfociti stimolati immunologicamente può essere responsabile dell'alta incidenza di linfomi associati ai disordini autoimmunitari e a rigetto di trapianti.

In generale, è ora ovvio che le interrelazioni che uniscono i virus, l'immunità e le malattie sono invero complesse. Sembra che la risposta immunitaria all'infezione virale possa avere sia effetti benefici che deleteri sull'ospite. Da una parte, essa può essere la piú importante o la sola arma contro l'infezione; dall'altra, può essere responsabile dei sintomi e anche degli effetti letali della malattia.

Un effetto secondario della interazione del complemento con i complessi antigene-anticorpo è l'attivazione di alcuni componenti del complemento (*in colore*) che aumentano la permeabilità dei vasi sanguigni e attraggono i globuli bianchi. A questo processo fa seguito una infiammazione, e gli enzimi liberati dai globuli bianchi possono danneggiare i tessuti non infettati, contribuendo a creare i dolorosi sintomi dell'infezione.

Fotografia di tessuto epatico di topo, in cui si vedono chiaramente numerosi globuli bianchi (*le cellule rotondeggianti colorate in scuro*). Le cellule sono state richiamate sul luogo di una lesione indotta da virus, probabilmente secondo lo schema della figura in alto: la cellula grande al centro del gruppo di cellule è stata infettata con *citomegalovirus*.

Le fonti energetiche del lavoro muscolare

La fonte diretta è l'adenosintrifosfato (ATP); indirette sono la combustione degli alimenti e la scissione del glicogeno. Le relazioni temporali di questi processi offrono validi spunti pratici

di Rodolfo Margaria

Il muscolo è un organo che produce energia meccanica a spese di energia chimica, a spese cioè della trasformazione di sostanze che si trovano nel muscolo preformate, o che vi giungono per mezzo della circolazione: è cioè un « motore ».

L'esame razionale di ogni motore, anche di quelli fabbricati dall'uomo, implica la misura esatta dell'energia meccanica prodotta, cioè del lavoro che il motore può compiere e dell'energia necessaria per azionarlo. L'analisi del lavoro meccanico è relativamente facile in fisiologia; viene compiuta generalmente servendosi di strumenti particolari detti ergometri.

La misura dell'energia spesa per azionare i motori fabbricati dall'uomo è facile: se si tratta di un motore elettrico basterà conoscere il voltaggio e l'intensità della corrente che alimenta il motore; se si tratta di un motore a combustione basterà misurare la quantità di carburante consumato, e cosí via. Nel caso dei muscoli la misura è molto piú difficile, tra l'altro perché varie sono le fonti energetiche che, direttamente o indirettamente, entrano in gioco in questo processo, e perché queste non possono essere misurate direttamente. Un bilancio energetico corretto può essere fatto soltanto in seguito a un'analisi quantitativa dettagliata di ciascuna di queste fonti energetiche e della relazione temporale del loro intervento rispetto alla produzione di lavoro motore.

L'unica fonte di energia che può essere sfruttata direttamente ai fini della produzione di lavoro nella contrazione muscolare appare essere quella contenuta nell'adenosintrifosfato (ATP) che, per distacco di una molecola di acido fosforico si trasforma in adenosindifosfato (ADP): ogni altra sorgente di energia non può essere sfruttata direttamente a questo fine. Poiché però l'ATP è contenuto nei muscoli in quantità molto piccola, dopo poche contrazioni il muscolo si esaurirebbe, e non sarebbe possibile un'attività continuata se l'ADP non venisse resintetizzato a spese di altri processi. Di questi il piú diretto è la scissione del creatinfosfato che, scindendosi in creatina e acido fosforico libera all'incirca la stessa quantità di energia necessaria per la resintesi dell'ATP.

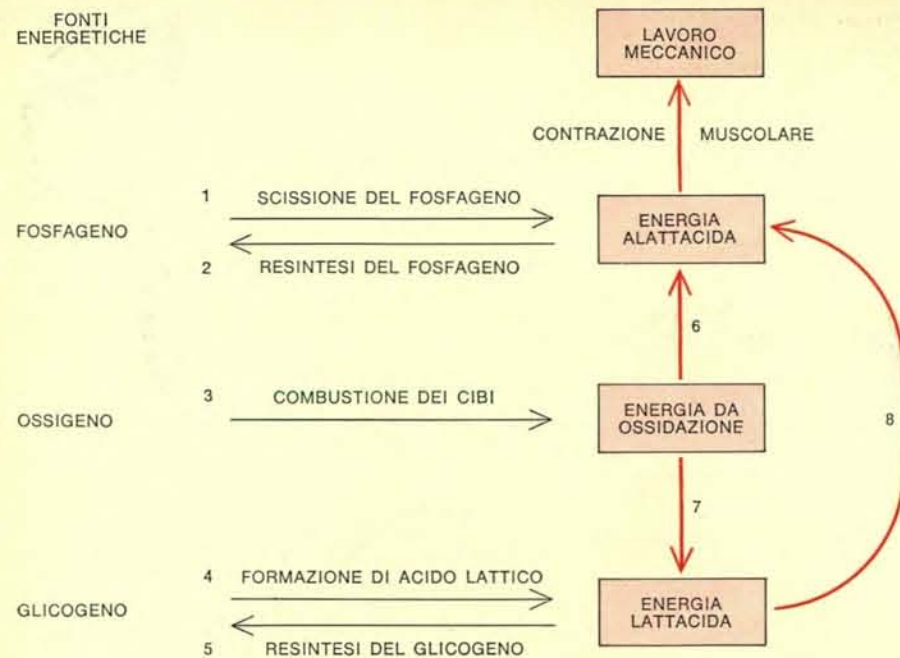
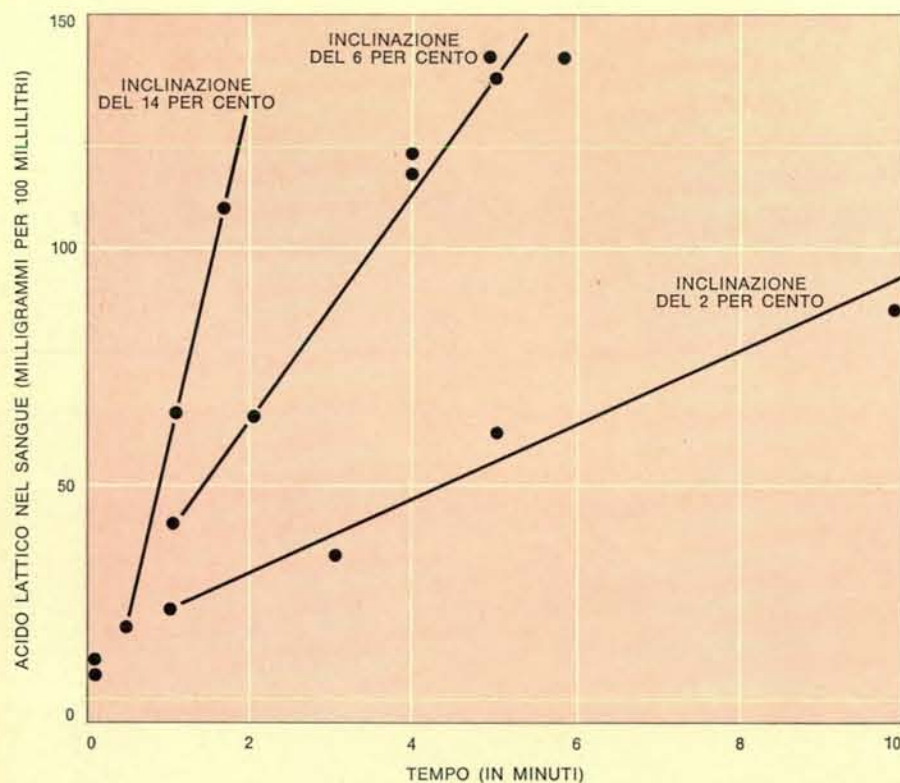
Il creatinfosfato si trova nei muscoli in concentrazione circa 3 volte superiore all'ATP: ma anche col suo contributo la riserva energetica dei muscoli non è tale da consentire un'attività prolungata.

Trattandosi di due reazioni in serie che hanno pressappoco lo stesso contenuto energetico, possiamo per semplicità denominare *fosfageno* tutte le sostanze contenenti fosfato altamente energetico e considerare la scissione del fosfageno la reazione energetica fondamentale della contrazione muscolare. Perché l'attività muscolare sia prolungata nel tempo occorre che il fosfageno, che è scisso durante la contrazione, venga resintetizzato, e perciò occorre che venga fornita energia al sistema: due sono le reazioni fondamentali che possono essere utilizzate a questo scopo, e cioè la combustione degli alimenti e la glicolisi.

L'entità delle combustioni può essere valutata dal consumo di ossigeno, quella della glicolisi dalla quantità di acido lattico che si è formato nell'organismo.

Le reazioni responsabili della produzione di energia nel lavoro muscolare sono rappresentate schematicamente nella figura in alto a pagina 42: sono esergoniche cioè liberano energia le reazioni *1*, *3* e *4* mentre sono endergoniche cioè assorbono energia le reazioni *2* e *5*. Alla conoscenza della energia liberata o assorbita in ognuna delle reazioni dello schema si può giungere attraverso dosaggi chimici nei muscoli del fosfageno per le reazioni *1* e *2*, o nel sangue e nei muscoli dell'acido lattico per le reazioni *4* e *5*, e attraverso il consumo di ossigeno per la reazione *3*. Ma per arrivare alla conoscenza, per ogni singolo processo energetico, della capacità – cioè della quantità totale di energia disponibile, e della potenza – cioè dell'energia trasformata nell'unità di tempo, occorre conoscere anche l'equivalente energetico di tutte le varie sostanze che prendono parte agli scambi energetici. Mentre era noto da tempo che l'equivalente calorico di 1 ml di ossigeno utilizzato nella combustione degli alimenti ammonta a circa 5 cal, l'equivalente calorico del fosfageno o dell'acido lattico formatosi dal glicogeno non era conosciuto con sicurezza, fino a poco tempo fa, perché determinazioni in vivo in condizioni fisiologiche non erano mai state eseguite: perciò non si poteva conoscere il costo energetico di prestazioni nelle quali una frazione preponderante di energia proveniva dalla scissione del fosfageno e dalla glicolisi, come per esempio la corsa dei 100 o dei 400 metri in gara.

È noto da tempo che nel lavoro sottomassimale non si ha produzione di acido lattico e che in ogni tipo di esercizio occorre un certo tempo (1-2 minuti) perché si giunga per il fosfageno a una condizione di equilibrio stazionario caratterizzato dal fatto che tanto è il fosfageno che si scinde quanto è quello che viene resintetizzato. Quindi, nel lavoro sottomassimale, che non richiede cioè un consumo di ossigeno per minuto superiore a quello massimo caratteristico per un singolo individuo e in condizioni di equilibrio, il dispendio

Un esercizio particolarmente faticoso come la corsa può portare l'atleta all'esaurimento delle sue risorse energetiche. Egli infatti consuma rapidamente tutto l'ATP contenuto nei muscoli che, per venire resintetizzato attraverso i processi di ossidazione e glicolisi, richiede un tempo relativamente lungo. Nella fotografia si vede Marcello Fiasconaro nelle ultime fasi della corsa dei 400 metri.

Rappresentazione schematica delle relazioni esistenti tra le varie fonti di energia muscolare. Il termine fosfageno indica genericamente un composto contenente fosforo altamente energetico come l'ATP e il creatinfosfato che si trovano nelle cellule. L'energia viene liberata dalla scissione delle molecole di fosfageno. Il fosfageno scisso viene subito resintetizzato; l'energia necessaria per la ricombinazione viene fornita da altre due fonti: la combustione dei cibi e la demolizione del glicogeno, un carboidrato che si accumula nei muscoli, fino ad acido lattico. Il glicogeno a sua volta viene resintetizzato a partire dall'acido lattico utilizzando l'energia proveniente dall'ossidazione.

Un esercizio sopramassimale consistente in una corsa in salita a velocità costante di 12 chilometri all'ora porta alla formazione di acido lattico, la cui concentrazione aumenta linearmente con la durata della prestazione. La produzione di acido lattico avviene più rapidamente quando la pendenza del terreno è maggiore e quindi l'esercizio più faticoso.

totale di energia sarà equivalente al consumo di ossigeno.

Per ciò che riguarda il contenuto energetico della glicolisi dobbiamo ricorrere a un tipo di esercizio sopramassimale, cioè superiore a quello che può essere sostenuto dalle reazioni di ossidazione, il cui limite è posto dall'ossigeno che può essere trasportato ai tessuti: in questo esercizio una condizione di equilibrio nel processo di scissione e resintesi del fosfageno è raggiunta molto presto, già entro poche decine di secondi: dopo questo tempo quindi il contributo di questa fonte energetica è nullo. Per di più, la resintesi del glicogeno dall'acido lattico avviene, come è noto da tempo, secondo un processo esponenziale molto lento, il cui tempo di semireazione è di 15 minuti; quando l'esercizio dura un tempo relativamente breve (1-5 minuti) questa reazione può quindi essere trascurata. Il dispendio energetico per minuto sarà perciò equivalente alla somma del massimo consumo di ossigeno per minuto e dell'acido lattico prodotto in un minuto.

Trattandosi di un lavoro sopramassimale, il massimo consumo di ossigeno ha un valore costante per un dato soggetto: anche se si aumenta la richiesta energetica, esso non può aumentare ulteriormente. In tali condizioni la produzione di acido lattico per minuto dipende direttamente dalla richiesta energetica totale.

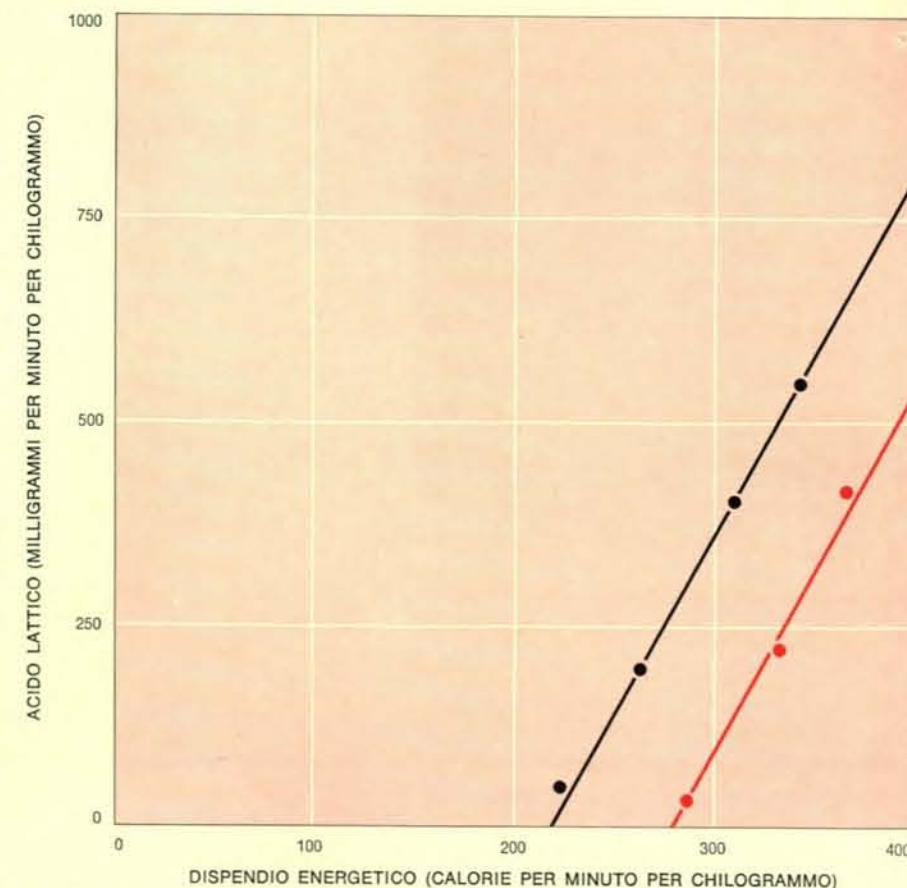
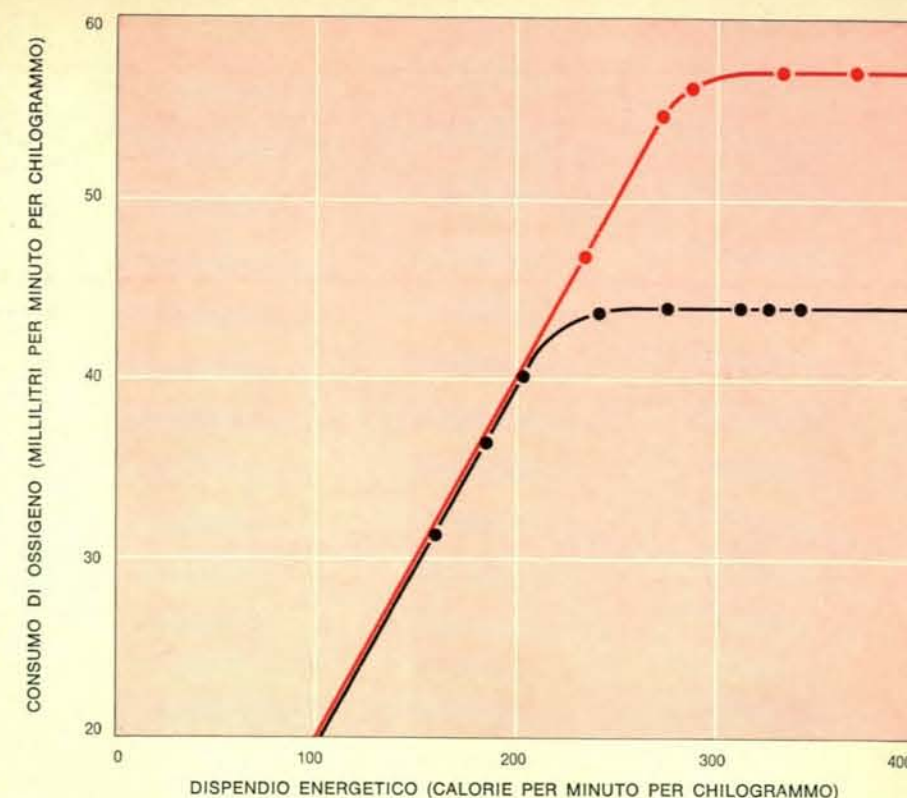
Questa ipotesi è stata riscontrata esatta al vaglio sperimentale. Si trattava di determinare la quantità di acido lattico liberatasi durante esercizi sopramassimali di varia intensità, che conducevano il soggetto a esaurimento in un tempo da 1 a 10 minuti. La raccolta di vari campioni di sangue per il dosaggio dell'acido lattico durante l'esercizio muscolare costituiva però un problema molto serio, sia perché non è possibile infilare un ago nella vena di un soggetto che, correndo, muove le braccia, sia perché la concentrazione dell'acido lattico nel sangue in quell'istante non sarebbe stata indicativa della quantità totale di acido lattico formatasi nei muscoli: si è ovviato a questa difficoltà facendo fermare il soggetto dopo un tempo prestabilito, di 1, 2, 3 minuti, a seconda dell'intensità dell'esercizio, e prelevando il sangue dopo circa 2-3 minuti dalla fine dell'esercizio, per permettere che l'acido lattico formato nei muscoli si distribuisse uniformemente in tutti i liquidi organici; soltanto quando è soddisfatta questa condizione, la concentrazione di acido lattico nel sangue è rappresentativa della quantità di acido lattico totale.

In una serie di esperimenti condotti con questo criterio a tre livelli di intensità lavorativa si sono potute costruire le rette della figura in basso nella pagina a fianco dalle quali appare che la concentrazione dell'acido lattico aumenta linearmente con la durata dell'esercizio. L'inclinazione delle rette, che è espressione della produzione di acido lattico nell'unità di tempo, aumenta con l'aumentare del carico lavorativo.

Dalla concentrazione dell'acido lattico nel sangue è possibile risalire ai grammi di acido lattico prodotti, quando si conosca il peso corporeo e il coefficiente di distribuzione dell'acido lattico nei vari organi e tessuti. Nella figura è indicato come varia la produzione di acido lattico in funzione della richiesta energetica: anche questa è una funzione rettilinea. La retta taglia l'ascissa a un valore di richiesta energetica corrispondente al massimo consumo di ossigeno. Nella figura in alto nella stessa pagina è indicato anche il consumo di ossigeno che a intensità di lavoro sottomassimale aumenta linearmente con la richiesta energetica, per giungere a un valore massimo costante quando la richiesta energetica supera un determinato valore.

I grafici delle figure a fianco mettono in evidenza che nel lavoro muscolare nello stato di equilibrio stazionario l'organismo normalmente sopperisce alla richiesta energetica con le sole reazioni ossidative, e soltanto quando queste non sono sufficienti (lavoro sopramassimale) interviene la glicolisi come meccanismo complementare di emergenza.

La quantità di energia che si libera nella formazione di 1 grammo di acido lattico è data dalla pendenza della retta; essa è indipendente dall'intensità dell'esercizio o dalla quantità totale di acido lattico che si forma, ed espressa in calorie ammonta a circa 230 calorie per grammo. È questo un valore termodinamico molto importante perché ottenuto non in vitro, ma nell'uomo in condizioni perfettamente fisiologiche, e cioè alla temperatura di 37 °C, al pH e alla forza ionica dei liquidi organici ecc.

Gli esperimenti da cui si sono ottenuti i dati delle figure a destra sono stati eseguiti su alcuni atleti, che avevano un massimo consumo di ossigeno notevolmente superiore a quello dei soggetti (studenti) ai quali si riferiscono gli esperimenti della figura in basso a pagina 42. Anche in essi si è osservato che acido lattico si forma soltanto nell'esercizio sopramassimale, e

I due grafici rappresentano il contributo energetico fornito dall'ossigeno (*in alto*) e dalla glicolisi (*in basso*) durante un esercizio. Nell'esercizio moderato quando la richiesta energetica indicata in ascissa come calorie per minuto per chilogrammo di peso è bassa, essa è soddisfatta semplicemente dall'ossidazione dei cibi corrispondente al consumo di ossigeno (misurato in millilitri per minuto per chilogrammo di peso corporeo). Quando la richiesta energetica supera il limite superiore del consumo di ossigeno la energia necessaria viene fornita dalla scissione del glicogeno in acido lattico (misurato in milligrammi per minuto per chilogrammo di peso corporeo). Gli atleti (*in colore*) presentano un più elevato consumo di ossigeno che ritarda la formazione di acido lattico, di conseguenza la linea corrispondente nel grafico è spostata più a destra.

che la retta che descrive la produzione di acido lattico in funzione della intensità dell'esercizio taglia l'ascissa allo stesso valore corrispondente al massimo consumo di ossigeno: questa linea è però piú spostata a destra. L'inclinazione della retta ha lo stesso valore, ed è ovvio che essa non debba variare, trattandosi di una costante termodinamica, che non può essere dipendente dalle capacità atletiche del soggetto.

In condizioni di anossia, cioè di carenza di ossigeno, quali possono essere create in laboratorio quando si respiri una miscela che contenga anziché il 21 per cento, il 12-14 per cento di ossigeno, oppure in alta montagna, il massimo consumo di ossigeno diminuisce. Corrispondentemente la retta che esprime la produzione di acido lattico in funzione della intensità dell'esercizio si sposta piú a sinistra, ma anche in questo caso la sua pendenza ha lo stesso valore che nei casi precedenti.

In altri esperimenti eseguiti su soggetti atleti e sedentari che avevano una capacità lavorativa notevolmente differente abbiamo mantenuto costante l'intensità del lavoro e cioè la richiesta energetica, e variato invece il massimo consumo di ossigeno. A questo scopo 12 soggetti che avevano un consumo di ossigeno che andava da un minimo di 27 a un massimo di 66 millilitri per chilogrammo per minuto è stato imposto un lavoro sopramassimale corrispondente a un consumo di ossigeno di circa 83 millilitri per chilogrammo per minuto (*si veda la figura a pagina 49*).

La produzione di acido lattico veniva misurata in tutti col metodo convenzionale, interrompendo cioè il lavoro in un dato momento e raccogliendo campioni di sangue per la determinazione dell'acido lattico dopo 3-4 minuti dalla fine del lavoro, quando si era stabilito un equilibrio di diffusione dell'acido lattico nei liquidi organici.

Si è osservato in questi esperimenti che l'acido lattico prodotto è sempre funzione lineare del tempo di prestazione: l'inclinazione delle rette, indice dell'acido lattico prodotto nell'unità di tempo, varia però da soggetto a soggetto, essendo maggiore nei soggetti aventi una minore potenza aerobica. Secondo quanto previsto, essendo costante la richiesta energetica, la produzione di acido lattico è funzione lineare del massimo consumo di ossigeno dei soggetti. Anche in questo caso la pendenza della retta non è altro che l'espressione dell'equivalente energetico dell'acido lattico, espresso in questo caso in millilitri di ossigeno anziché in calorie.

In questa serie di esperimenti l'equivalente energetico dell'acido lattico è risultato avere un valore di 245 calorie per grammo, molto prossimo a quello ottenuto nella prima serie di esperimenti, nei quali il massimo consumo di ossigeno era mantenuto costante e variava invece la richiesta energetica.

Stabilito il valore del coefficiente calorico dell'acido lattico è facile risalire alla capacità di questo meccanismo, essendo noto che il massimo aumento di concentrazione dell'acido lattico, per effetto dell'esercizio muscolare, ammonta a 1,5 grammi per litro di sangue circa, corrispondente a circa 1,12 grammi per chilogrammo di peso corporeo. La massima quantità di energia ottenibile dalla formazione di acido lattico ammonta quindi a circa 260 calorie per chilogrammo di peso corporeo, corrispondente cioè a un consumo di ossigeno di circa 52 millilitri per chilogrammo.

Come il consumo di ossigeno aumenta con l'aumentare della richiesta energetica fino a un determinato limite, che è presumibilmente determinato dalla quantità di ossigeno che può essere trasportata col sangue ai muscoli attivi, cosí anche la quantità di acido lattico che può essere prodotta nell'unità di tempo ha un limite massimo, che è dato presumibilmente dalla massima velocità delle reazioni chimiche della catena della glicolisi. Si è visto infatti che quando si aumenta l'intensità del lavoro imposto ai soggetti oltre un determinato valore, l'acido lattico prodotto nell'unità di tempo non aumenta piú.

Questa massima produzione di acido lattico (*si veda la figura a fianco*) è di circa 1,7 grammi per minuto per chilogrammo e corrisponde a una potenza di circa 365 calorie per minuto per chilogrammo.

L'intensità dell'esercizio in questi esperimenti era cosí elevata da condurre all'esaurimento in un tempo variabile da 5 a 35 secondi ed è interessante notare che, malgrado la gravità dell'esercizio, non si formava acido lattico se il lavoro era di breve durata; compariva cioè soltanto dopo un certo tempo dall'inizio del lavoro, 15 secondi circa, nel lavoro meno pesante che conduceva all'esaurimento in 35 secondi (*si vedano le figure nella pagina a fronte*). In questa prima fase l'energia richiesta dall'esercizio non poteva essere attribuita al meccanismo glicolitico e anche il meccanismo ossidativo, che è piuttosto lento a mettersi in moto, non poteva contribuire al bilancio energetico che in misura molto ridotta: quasi tutta l'energia doveva quindi provenire dalla scissione del fosfageno.

Esercizi particolarmente faticosi portano la velocità della produzione di acido lattico al limite superiore, al di sopra del quale, anche l'aumento di intensità dell'esercizio non ha piú influenza, come dimostrano le linee parallele del grafico. L'esercizio consiste nella corsa su nastro alla velocità di 18 chilometri all'ora con inclinazione variabile dal 10 al 25 per cento. Ogni retta termina in corrispondenza del punto di esaurimento (*in colore*) raggiunto tanto piú rapidamente quanto piú l'esercizio è faticoso.

Come aumenti il consumo di ossigeno all'inizio di un lavoro strenuo sopramassimale era noto da esperimenti precedenti nei quali analisi di aria espirata venivano eseguite in rapida successione all'inizio del lavoro: si era riscontrato che il consumo di ossigeno aumenta secondo un processo esponenziale, la cui velocità era proporzionale alla richiesta energetica: il tempo di semireazione di questo processo è di circa 30 secondi. Da questi dati è possibile calcolare il contributo energetico delle ossidazioni nei primi istanti del lavoro negli esperimenti riassunti nel grafico della pagina a fianco e portare una opportuna correzione.

Sottraendo cioè dalla quantità di energia totale richiesta nell'esercizio, che è nota, l'energia ossidativa e l'energia dovuta all'acido lattico formatosi, è possibile ottenere la quantità di energia imputabile alla scissione del fosfageno. Si è calcolato che la quantità massima teorica di energia per chilogrammo di peso ottenibile dal fosfageno equivale a 200 calorie. Poiché la resintesi del fosfageno a spese delle ossidazioni ha luogo con un rendimento di circa il 55 per cento, tale energia corrisponde a un valore di $200 \times 0,55 = 110$ cal/kg di energia ottenibile dalla scissione del fosfageno; assegnando al fosfageno un equivalente calorico di 11 calorie per millimole, il fosfageno scisso ammonta a circa 10 millimoli per chilogrammo di peso corporeo. Poiché i muscoli rappresentano il 40-50 per cento del peso corporeo, tale valore corrisponde a un contenuto in fosfageno nei muscoli di circa 22-25 millimoli per chilogrammo di muscolo fresco, un valore molto prossimo a quello ottenuto dalle analisi chimiche del muscolo umano prelevato a mezzo di biopsie da autori scandinavi.

I bilanci energetici dei quattro espe-

I quattro grafici a destra mostrano i contributi forniti dalle tre fonti energetiche durante determinati esercizi sopramassimali. Gli esercizi consistevano nella corsa fino a esaurimento su nastro trasportatore alla velocità di 18 chilometri all'ora e con inclinazioni variabili tra il 10 e il 25 per cento. Il contributo energetico fornito dal processo alattacido (*area tratteggiata*) deriva dalla scissione del fosfageno e costituisce la prima e piú importante fonte di energia. L'ossidazione del cibo (*area colorata*) fornisce un'energia che cresce esponenzialmente a partire dall'inizio dell'esercizio, ma il suo contributo è abbastanza scarso soprattutto nei primi secondi. La scissione del glicogeno in acido lattico (*area grigia*) fornisce la restante energia.

rimenti rappresentati nelle rette della figura a pagina 44 sono stati graficamente rappresentati nelle figure di pagina 45.

Nelle considerazioni sopraesposte sulla formazione dell'acido lattico e sul bilancio energetico del muscolo si è partiti dal presupposto che tutto l'acido lattico venga prodotto dai muscoli nella fase di attività e che non ne venga affatto prodotto nella fase di ristoro. In base a questo presupposto si può calcolare il fosfageno che si scinde nella attività muscolare: si è cosí riscontrato che la massima quantità di fosfageno che si scinde nel lavoro ammonta a solo la metà del valore massimo teorico, e cioè a circa 100 calorie per chilogrammo di peso corporeo, corrispondenti a 20 millilitri di ossigeno.

Non è possibile misurare istante per istante la produzione di acido lattico durante l'esercizio e nel susseguente periodo di ristoro. È noto però da esperimenti sull'uomo nel lavoro muscolare strenuo e sul muscolo isolato che la produzione di acido lattico non è limitata al periodo di attività muscolare, ma si prolunga anche nelle prime fasi del ristoro; questa produzione di acido lattico, come del resto anche quella che si forma nel periodo di attività, conformemente alle reazioni *2* e *4* dello schema in alto a pagina 42 non ha altro significato che quello di trasformare il debito di ossigeno alattacido, dovuto cioè alla scissione del fosfageno, in debito lattacido. Si tratta di un vero e proprio « ristoro anaerobico » un processo che è stato descritto e studiato quantitativamente fin dal 1936 da G. Moruzzi e da me sul muscolo isolato di rana.

Certamente durante l'esercizio strenuo una quota di fosfageno muscolare molto maggiore di quella che risulta dal bilancio finale viene scissa, forse tutta o quasi: e alla fine del lavoro una notevole frazione, circa la metà, verrebbe resintetizzata a spese dei processi glicolitici, mentre l'altra metà verrebbe resintetizzata a spese delle ossidazioni organiche, e apparirebbe come debito alattacido vero e proprio.

Lo schema della figura a pagina 42 in alto ha anche il vantaggio, oltre quello di darci una descrizione delle trasformazioni energetiche che hanno luogo nel muscolo che lavora, di illustrarci il significato e le caratteristiche anche temporali della contrazione e del pagamento del debito di ossigeno. Come è noto, il termine « debito di ossigeno » è stato introdotto da Hill verso il 1920 per indicare quella quantità di energia spesa che all'inizio dell'esercizio proviene non da reazioni ossidative, ma anaerobiche, e che viene restituita al muscolo alla fine del lavoro ad opera di reazioni ossidative: questa quantità di energia viene sempre espressa in millilitri di ossigeno.

Alcuni intervalli di riposo tra una corsa faticosa e un'altra possono ridurre notevolmente la produzione di acido lattico e aumentare la lunghezza totale del percorso. L'esercizio consiste nella corsa su nastro trasportatore alla velocità di 18 chilometri all'ora per dieci secondi seguiti da un periodo di riposo. Quando il riposo è di 10 secondi (*a*) il corridore può ripetere l'esercizio dieci volte prima di arrivare all'esaurimento. Quando il riposo è di 20 secondi (*b*) l'esercizio può essere ripetuto 20 volte. Con un periodo di 30 secondi tra una corsa e l'altra (*c*), il livello di acido lattico rimane costante e l'esercizio può essere continuato all'infinito. Gli atleti possono dunque allenarsi meglio e piú a lungo alternando agli esercizi opportuni periodi di riposo.

La reazione *1* è espressione anche, oltre che della potenza muscolare, della contrazione di un debito di ossigeno alattacido. Poiché la reazione *2* ha luogo con un certo ritardo rispetto alla prima, esiste sempre nel muscolo in attività a equilibrio una certa quantità di fosfageno scisso, che è l'espressione della entità del debito di ossigeno alattacido; questa quantità sarà tanto piú grande quanto piú elevata è la velocità della reazione *1*, e cioè quanto piú grande è la potenza sviluppata dai muscoli. La reazione a catena *3-6-2*, è l'espressione del pagamento del debito di ossigeno alattacido. Le reazioni *4-8-2* sono invece l'espressione della contrazione del debito di ossigeno lattacido, e questo, come appare dallo schema, ha sempre il significato del pagamento del debito alattacido: la velocità di questa reazione, espressione della potenza del meccanismo lattacido, ha un limite, che è dato dalla velocità massima della reazione *4*. La reazione *3-7-5* è l'espressione del pagamento del debito di ossigeno lattacido e questa, come si è detto, è una reazione molto lenta, il limite di velocità essendo posto dalla lentezza della reazione *5*.

Risulta dai dati della figura a pagina 44 che quanto maggiore è l'intensità dell'esercizio, e cioè la potenza sviluppata, tanto minore è il tempo per cui quell'esercizio può essere mantenuto.

Si è riscontrato che la massima potenza che si misura sperimentalmente è quella che si ottiene estrapolando la curva della potenza in funzione della durata dell'esercizio, non a zero, ma a un valore di durata dell'esercizio di 5 secondi. Questo sta a indicare che anche la potenza muscolare, dovuta alla scissione del fosfageno, come la potenza aerobica e la potenza dovuta al meccanismo lattacido, ha delle limitazioni, dovute presumibilmente alla velocità delle reazioni chimiche implicate nella scissione del fosfageno; questa massima potenza potrà essere mantenuta solo per pochissimi secondi, circa 4-5, per il rapido esaurirsi della scorta di fosfageno dei muscoli: per questo breve tempo, si mantiene costante dopodiché decresce: e anche quando si richieda un esercizio di minor durata, non aumenta ulteriormente.

La misura della potenza dovuta al meccanismo alattacido può essere fatta con un metodo molto semplice. Quando si salga una scala alla massima velocità a due gradini per volta, la velocità, dopo una breve fase di accelerazione iniziale, che dura meno di 2 secondi, si mantiene costante per quattro secondi, dopodiché decresce progressivamente. Poiché in questo esercizio il lavoro esterno compiuto tra il secondo e il quarto secondo consiste quasi esclusivamente nel sollevamento del corpo, essendo solo una frazione trascurabile impiegata a sostenere le variazioni di velocità a ogni passo, la misura della componente verticale della velocità in metri al secondo, nel periodo di velocità costante, darà anche l'indicazione della potenza meccanica sviluppata espressa in chilogrammetri per secondo e per chilogrammo di peso corporeo. Occorre naturalmente conoscere l'altezza dei gradini, e disporre di un cronometro elettronico a cellule fotoelettriche, sensibile a un centesimo di secondo per il calcolo della velocità verticale.

Poiché da dati sperimentali ottenuti nella corsa in salita il rendimento di questo esercizio è circa 0,25, basterà dividere il lavoro meccanico per questo valore per risalire al valore di consumo di energia. Questa prova richiede pochissimo tempo di esecuzione, è bene accetta dai soggetti poiché non conduce all'esaurimento e può essere utilmente impiegata su larga scala, per esempio nello studio di popolazioni: essa ha l'inconveniente che misura soprattutto la potenza dei muscoli degli arti inferiori, e di questo bisogna tenere conto quando si voglia analizzare la capacità lavorativa di individui che debbono svolgere una determinata mansione con le braccia anziché con le gambe.

La massima potenza muscolare dovuta alla scissione del fosfageno, cosí misurata, appare essere all'incirca tre volte maggiore della massima potenza dovuta alle reazioni ossidative, ammonta cioè a circa 800 calorie per chilogrammo di peso e per minuto corrispondente a un consumo di ossigeno di 160 millilitri per chilogrammo per minuto.

Anche la massima potenza aerobica, espressione della massima velocità della reazione *3*, può essere misurata facilmente, oltre che direttamente, con un metodo indiretto piú pratico, che si basa sui presupposti, riscontrati sperimentalmente validi con sufficiente approssimazione, a) che la frequenza cardiaca aumenta linearmente col consumo di ossigeno e b) che la massima frequenza cardiaca ha un valore definito in funzione dell'età.

Mentre la potenza e la capacità sia del meccanismo alattacido che del meccanismo ossidativo sono costanti e caratteristiche per un dato individuo, la capacità e la potenza del meccanismo lattacido sono molto variabili, poiché dipendono, come è ovvio, dal contenuto in glicogeno dei muscoli e questo varia in dipendenza dallo stato di nutrizione del soggetto al momento della prova. Il rilievo di questo carattere non è perciò, a mio avviso, molto interessante, poiché non è una caratteristica definita e costante dell'individuo. Ovviamente è importante che un individuo che si presenta a una prova impegnativa, quale è per esempio una competizione sportiva, sia in buone condizioni di nutrizione, almeno per ciò che riguarda il contenuto in glicogeno dei muscoli.

La massima quantità di acido lattico prodotto nell'esercizio muscolare ammonta, come si è detto, a 1,12 grammi per chilogrammo di peso corporeo. Poiché il muscolo contiene, secondo i dati pubblicati da Karlsson e Saltin (1970) circa 14 grammi di glicogeno per chilogrammo, nel presupposto che le masse muscolari costituiscano il 40 per cento del peso corporeo, soltanto un quinto del glicogeno muscolare avrebbe la possibilità di essere trasformato in acido lattico per fornire energia nel lavoro anaerobico. Questi 1,12 grammi di acido lattico corrispondono a 260 calorie o a 52 millilitri di ossigeno, una quantità cioè pressappoco equivalente al massimo consumo di ossigeno in un minuto.

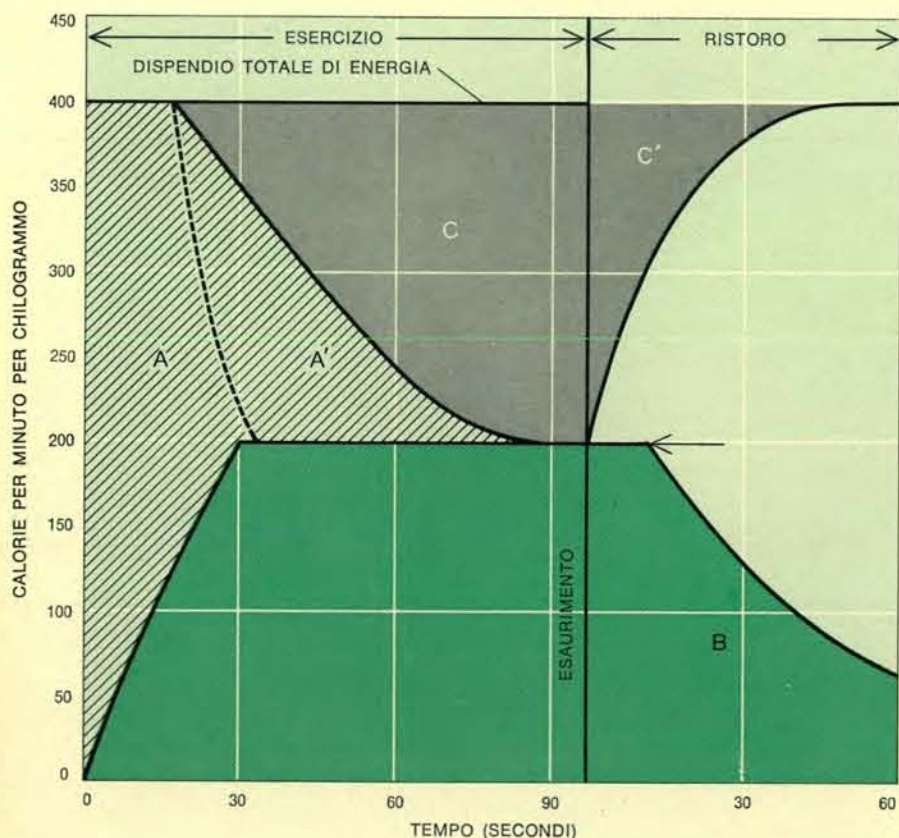
In pratica per l'organismo è conveniente limitarsi alla contrazione di un debito alattacido e non incorrere in un debito lattacido quando le condizioni dell'esercizio lo permettano. Infatti il pagamento del debito lattacido è molto lento e non è ancora completo dopo un'ora di ristoro. Per di piú, l'acido lattico provoca una condizione di acidosi che si manifesta con una sintomatologia oggettiva e soggettiva imponente, consistente tra l'altro in una forte diminuzione della possibilità di prestazione, e in una sensazione grave di malessere. Il debito di ossigeno alattacido invece non implica alcuna alterazione dell'equilibrio acido-base del sangue e dei liquidi organici, e viene pagato completamente in pochissimi minuti.

Tenendo conto del fatto che la produzione di acido lattico entra in gioco con un certo ritardo nel lavoro muscolare si può predire che si può compiere un lavoro sopramassimale anche di lunga durata senza che si abbia alcuna produzione di acido lattico, purché il lavoro sia intermittente. Gli intervalli di lavoro devono essere sufficientemente brevi, tali che non si abbia produzione di acido lattico e il periodo di ristoro sufficientemente lungo da permettere che abbia luogo il ristoro anaerobico, che, d'altra parte, consistendo nel pagamento del debito alattacido di ossigeno è, come si è detto, un processo rapido.

Questo è quanto si verifica sperimentalmente: dei soggetti sono stati invitati a correre su un ergometro trasportatore alla velocità di 18 chilometri all'ora e con una pendenza del 15 per cento, un esercizio che avrebbe condotto all'esaurimento in un tempo variabile, a seconda dei soggetti, da 32 a 38 secondi; l'intensità di questo esercizio è paragonabile a quella di un atleta nella corsa dei 400 metri in gara. La durata della corsa era limitata a soli 10 secondi, seguita da un periodo di riposo che nelle diverse prove era di 10, 20 o 30 secondi, dopodiché il ciclo si ripeteva fino a esaurimento del soggetto (*si veda la figura nella pagina a fronte*).

Si è osservato in questi esperimenti che se il periodo di riposo era di 10 secondi il soggetto poteva compiere circa 10 periodi di corsa, corrispondenti a un tempo totale di esercizio di 100 secondi, mentre l'acido lattico nel sangue saliva a 1,15 grammi per litro. Se invece il periodo di riposo era di 20 secondi l'esaurimento non sopravveniva che dopo venti cicli circa, corrispondenti a un tempo totale di esercizio di 200 secondi e la concentrazione dell'acido lattico nel sangue era di 0,8 grammi per litro. Se il periodo di riposo era di 30 secondi, dopo un breve iniziale aumento, la concentrazione dell'acido lattico nel sangue rimaneva costante a circa 0,2 grammi per litro, mentre il lavoro poteva essere protratto indefinitamente. La lieve salita iniziale dell'acido lattico è dovuta al lento mettersi in moto del meccanismo ossidativo, e quindi a una condizione di relativa anaerobiosi nella quale si viene a trovare il soggetto all'inizio dell'esercizio: questa condizione si verifica anche quando l'intensità del lavoro non è sopramassimale.

È ovvio che la somma di lavoro che si può compiere in un certo lasso di tempo è molto maggiore se il lavoro è intermittente, che non se il lavoro è protratto fino all'esaurimento in un'unica prova. Cosí per esempio se un quattrocentometrista si vuole allenare percorrendo i quattrocento metri alla massima velocità si troverà alla fine della prova in condizioni di grave acidosi,

Quando i muscoli ricavano energia dalla scissione del fosfageno contraggono un debito di ossigeno alattacido *(area tratteggiata A)*; **contemporaneamente si va attivando il meccanismo di ossidazione. Nell'esercizio moderato che non porta alla formazione di acido lattico, il consumo di ossigeno raggiunge rapidamente il limite superiore di 40 millilitri (equivalenti a 200 calorie) per minuto per chilogrammo di peso corporeo e fornisce tutta l'energia richiesta dai muscoli. Alla fine dell'esercizio viene consumato ulteriore ossigeno** *(area colorata B)* **per pagare il debito di ossigeno alattacido** *(A)*.

Quando l'esercizio è sopramassimale e porta all'esaurimento vengono contratti debiti lattacido e alattacido. Il debito d'ossigeno alattacido *(aree tratteggiate A e A')* **si forma quando i muscoli ricavano energia dalla scissione del fosfageno, mentre l'energia proveniente dal consumo di ossigeno** *(area in colore)* **sta raggiungendo il suo limite superiore** *(freccia)*. **Il debito di ossigeno lattacido** *(aree grige C e C')* **viene contratto quando l'energia viene ottenuta dall'acido lattico. Durante il periodo di ristoro il consumo di ossigeno rimane a livelli elevati per 15-20 secondi e quindi cala rapidamente. La energia ricavata dal consumo di ossigeno** *(area B)* **serve a pagare la maggior parte** *(area A)* **del debito di ossigeno alattacido. La produzione di acido lattico continua** *(C')* **dopo l'esercizio per fornire l'energia necessaria a pagare la frazione A' del debito alattacido.**

che potrà essere rimossa soltanto dopo un periodo di ristoro di almeno un'ora e mezzo; in una mattinata (3 ore) egli potrà ripetere la prova due volte e percorrere complessivamente 800 metri. Se invece egli corre alla stessa velocità per cento metri soltanto, e dopo un intervallo di 30 secondi, ripete la prova, e cosí via, egli potrà senza avere alcun aumento dell'acido lattico nel sangue, quindi senza venirsi a trovare in condizioni di acidosi, compiere in tre ore 270 intervalli di corsa di 100 metri ognuno, totalizzando quindi un percorso di 27 chilometri, corrispondente a un lavoro totale piú di 30 volte superiore al massimo lavoro che avrebbe potuto compiere se questo fosse stato continuo.

La possibilità di poter ricorrere a questo sistema per aumentare la possibilità di lavoro totale svolto non si applica ovviamente soltanto allo sport, ma anche al lavoro industriale o produttivo.

Si è accennato precedentemente alla difficoltà pratica di misurare la quantità di acido lattico prodotto nel corso dell'esercizio muscolare, perché, affinché l'analisi dell'acido lattico nel sangue sia significativa, occorre attendere un tempo sufficiente dopo la fine del lavoro per permettere che l'acido lattico diffonda dall'interno delle cellule muscolari nei liquidi organici e vi si distribuisca uniformemente. I dati delle figure in basso a pagina 42 e a pagina 44, calcolati partendo dal presupposto che tutto l'acido lattico che si viene a trovare alla fine dell'esercizio sia stato prodotto durante l'esercizio stesso, si riferiscono dunque al bilancio finale, ma non sono necessariamente rappresentativi dell'andamento temporale dei vari processi chimici ed energetici che hanno luogo nel corso dell'esercizio.

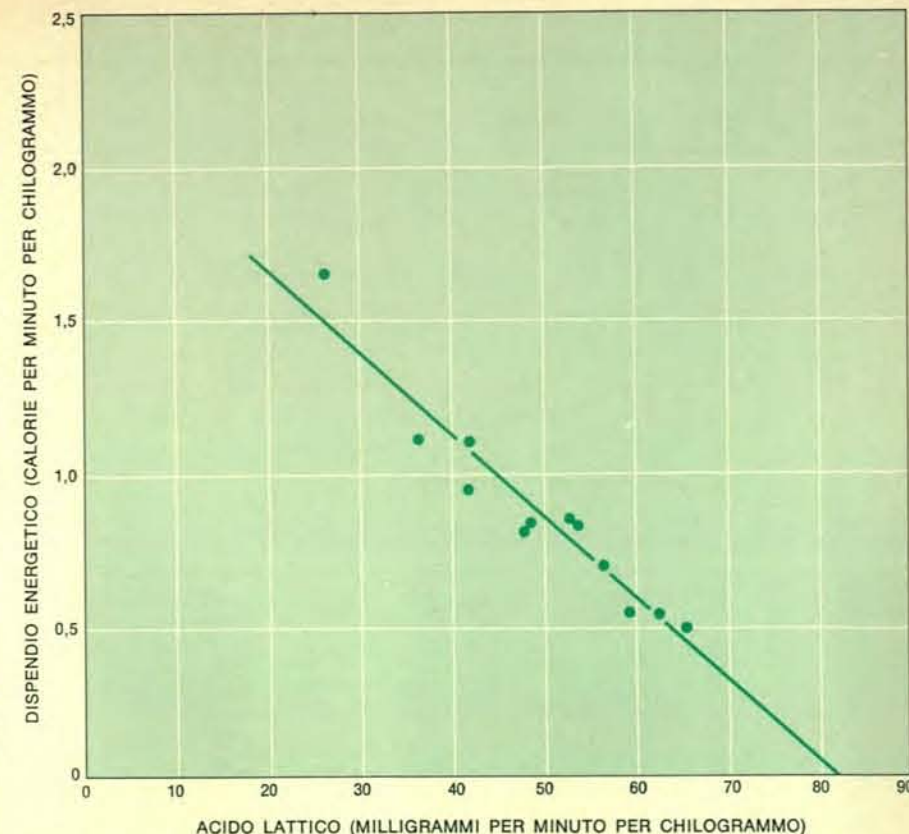
In effetti parte dell'acido lattico è certamente prodotto dopo la cessazione del lavoro, nel primo periodo del ristoro, come è dimostrato dal fatto che nel lavoro sopramassimale che conduce all'esaurimento in 1-2 minuti, la concentrazione dell'acido lattico nel sangue non decresce immediatamente alla fine del lavoro, ma si mantiene elevata per 4-5 minuti; e soprattutto dal fatto che in tale tipo di esercizio, malgrado che alla sua cessazione la ventilazione polmonare scenda bruscamente, il quoziente respiratorio sale a valori molto elevati, fino a 2 e oltre, indice verosimilmente del versamento di acidi in circolo dopo la cessazione dell'attività muscolare. D'altra parte un ritardo nella produzione dell'acido lattico rispetto agli eventi meccanici era stato già dimostrato nel muscolo isolato di rana da Embden nel 1920.

Che la produzione di acido lattico sia ritardata rispetto agli eventi meccanici, che devono necessariamente coincidere con la scissione dell'ATP nel muscolo, risulta ovvio quando si consideri che le reazioni *4-8-2* nello schema della figura in alto a pagina 42 devono essere successive alla reazione *1*: queste reazioni in altre parole non hanno altro significato, come si è già accennato precedentemente, che quello della trasformazione del debito alattacido in debito lattacido. Stando cosí le cose nel bilancio finale il debito alattacido apparirà minore di quello che è stato effettivamente contratto.

Questo spiega perché il debito alattacido da noi riscontrato nel bilancio finale non ammonti che a 20 millilitri di ossigeno, corrispondenti, come è stato rilevato precedentemente, alla scissione di circa la metà del fosfageno contenuto nei muscoli. Presumibilmente tutto il fosfageno è disponibile per la contrazione del debito di ossigeno alattacido: la metà di questo debito viene pagata direttamente, e si manifesta come aumento del consumo di ossigeno nei primissimi minuti del ristoro, mentre l'altra metà viene pagata indirettamente, tramite la contrazione di un debito lattacido.

Come è noto da tempo, il consumo di ossigeno decresce istantaneamente alla fine del lavoro secondo il noto andamento esponenziale: ma la curva del consumo di ossigeno nel ristoro è sempre stata eseguita finora soltanto a seguito del lavoro aerobico. Recentemente l'analisi della curva del consumo di ossigeno nel ristoro è stata eseguita in seguito a un lavoro sopramassimale molto intenso, tale da condurre all'esaurimento in 15-20 secondi, e si è osservato che la curva del consumo di ossigeno è allora molto piú complessa: esso si mantiene elevato per i primi 15-20 secondi circa del ristoro, e soltanto in seguito decresce con la consueta modalità. Il debito di ossigeno misurato in queste condizioni è risultato essere notevolmente superiore, circa il doppio, del debito di ossigeno che si ha in seguito al massimo lavoro muscolare aerobico.

Si è avuta cosí la conferma che tutto il fosfageno contenuto nei muscoli può essere utilizzato per la contrazione del debito di ossigeno alattacido. Parte di questo debito, alla fine di un lavoro strenuo, viene trasformata in debito lattacido.

Questi risultati permettono di tracciare un quadro almeno approssimativo degli eventi che hanno luogo nel

Facendo eseguire a vari soggetti atleti e non atleti un esercizio della stessa intensità e quindi mantenendo costante la richiesta energetica, si è potuto mettere in relazione l'aumento della concentrazione di acido lattico nel sangue con la massima potenza aerobica variabile secondo i soggetti (espressa come consumo di ossigeno).

muscolo durante l'attività e il ristoro. Nelle figure a pagina 48 è indicato come aumenta il consumo di ossigeno all'inizio del lavoro e come decresce nel periodo di ristoro quando l'esercizio sia massimale, corrispondente cioè al massimo consumo di ossigeno; in queste condizioni non si forma acido lattico e il debito di ossigeno è soltanto alattacido. Ma se la richiesta energetica è due volte quella corrispondente al massimo consumo di ossigeno, il consumo di ossigeno sale con una velocità doppia, per mantenersi costante quando si è raggiunto il valore limite che non può essere superato. All'inizio la richiesta energetica è soddisfatta sostanzialmente dalla scissione del fosfageno, che si esprime come contrazione del debito alattacido. Acido lattico incomincia a formarsi soltanto quando una notevole quantità del fosfageno muscolare è scisso; quando tutto il fosfageno è scisso la produzione di acido lattico costituisce il solo meccanismo responsabile per la quota sopramassimale di richiesta energetica. Nel ristoro la produzione di acido lattico continua, contribuendo cosí al pagamento del debito alattacido contratto durante l'esercizio. Il rimanente debito alattacido è pagato direttamente dalle ossidazioni. Nella figura non è rappresentato il pagamento del debito lattacido, poiché questo è un processo lento che dura oltre un'ora.

Le trasformazioni chimiche ed energetiche che hanno luogo nel muscolo e che risultavano ancora molto poco chiare solo pochissimi anni or sono, oggi sono sufficientemente delucidate da rendere ragione, almeno nelle grandi linee, della possibilità da parte dell'organismo umano o animale di compiere lavoro. Da questa conoscenza è possibile anche impostare un piano di lavoro tale da fornire la migliore prestazione energetica, sia che si tratti di sviluppare la massima potenza meccanica, o di compiere la massima quantità di lavoro in un determinato tempo. Queste conoscenze possono trovare larghe applicazioni non solo nell'ambito delle prestazioni sportive, ma, ciò che è piú importante dal punto di vista sociale, in campo ergonomico, per la migliore utilizzazione dell'energia disponibile da parte del lavoratore, a vantaggio del rendimento lavorativo, e cioè delle capacità produttive e soprattutto delle condizioni di benessere e di salute di chiunque sia impegnato in un lavoro professionale prolungato che richieda produzione di lavoro meccanico.

Il giudaismo al tempo di Cristo

Il giudaismo è tradizionalmente considerato un complesso unificato di credenze. Si ritiene ora invece che tra il 200 a.C. e il 100 d.C. esistessero diverse espressioni religiose, alcune delle quali addirittura pagane

di Michael E. Stone

Tutta una serie di scoperte effettuate negli ultimi decenni ha portato a mutare in modo sorprendente la moderna concezione del giudaismo e del cristianesimo durante il periodo compreso tra il 200 a.C. e il 100 d.C. Con le loro ricerche, studiosi di numerose discipline hanno svelato insospettate dimensioni nella diversità dell'espressione religiosa dei giudei. La creatività del giudaismo nel periodo a cavallo della nascita di Cristo può esserne la principale caratteristica.

Con la conclusione, nella Bibbia ebraica, dei grandi racconti storici intorno al 400 a.C., nella nostra conoscenza del giudaismo c'è un vuoto relativo ai due secoli successivi. Sebbene una parte della letteratura biblica venisse organizzata e forse addirittura scritta durante questi secoli, non si sa quasi nulla sullo sviluppo del giudaismo in tale periodo. Si trattava comunque di un tempo di grandi mutamenti militari e politici in tutto il Medio Oriente: Alessandro Magno aveva conquistato l'impero persiano e l'ellenizzazione delle antiche culture indigene di questa area era ormai in atto. L'età ellenistica fu un periodo caratterizzato da una insolita varietà di religioni, di alte teosofie e arti magiche, di filosofie e misteri. Quando si sollevò il velo sul giudaismo dopo il 200 a.C., il mondo ebraico era molto diverso da quello delineato nell'ultima letteratura biblica. A cominciare, nel 167 a.C., con la rivolta dei maccabei contro i successori seleucidi di Alessandro, ne esistono numerose testimonianze letterarie.

Il giudaismo rivelò una fecondità di espressione tipica di altre religioni ellenistiche. Eppure, di tutte le molteplici forme espressive della religione giudaica che videro la luce, solo due sono sopravvissute e fioriscono tuttora: la forma rabbinica, da cui discende tutto il giudaismo moderno, e la forma nota come cristianesimo. Superstite di marginale importanza è anche il samaritanesimo, mantenuto in vita da un esiguo numero di adepti a Shechem, nella Palestina centrale, che continuano i loro culti sul sacro Monte Gerizim dove, in tempi antichi, sorgeva il loro tempio. Se noi dipendessimo solo dagli scritti tramandatici all'interno della tradizione giudaica, sapremmo molto poco del giudaismo al tempo di Cristo. Ogni informazione su questi gruppi, che il giudaismo rabbinico osteggiava, è stata sistematicamente ignorata o soppressa negli scritti rabbinici. Non sapremmo nulla delle fortune politiche della dinastia degli asmonei, che si insediarono dopo la rivolta dei maccabei, né dei ribelli assediati dai romani a Masada. Sarebbero andate perdute molte delle nostre conoscenze sui sadducei e sugli esseni dei deserti della Palestina. Le notizie su costoro e sul giudaismo della Diaspora ellenistica – l'importante insediamento ebreo situato oltre i confini della Palestina – sono state trasmesse fino a noi al di fuori della tradizione giudea. Poiché, nello spettro del giudaismo, il cristianesimo affondava le sue radici in una parte diversa, preferenze e pregiudizi di quest'ultimo erano diversi, e molto diverso fu il tipo di scritti giudaici che venne conservato. D'altra parte solo ora sono stati pienamente assimilati i documenti dovuti a due importantissimi ritrovamenti – i rotoli del Mar Morto, rinvenuti a Qumran, e i manoscritti gnostici di Nag Hammadi, in Egitto.

Il primo indizio della varietà delle espressioni religiose giudaiche nel periodo compreso tra il 200 a.C. e il 100 d.C. sorse forse con la traduzione nelle lingue europee dei libri giudaici conservati solo dalla chiesa etiopica e da altre chiese cristiane orientali. Tra i piú importanti erano il libro di Enoc e il libro dei Giubilei, ambedue tradotti dall'etiopico nel XIX secolo. Era evidente che il libro di Enoc era stato la fonte della lettera di Giuda nel Nuovo Testamento e di altri scritti cristiani piú antichi. Si confrontino, per esempio, i due brani seguenti, il primo tratto dal libro di Enoc e il secondo dalla lettera di Giuda: « Ed ecco! Egli giunge con le miriadi delle sue sante schiere per eseguir giudizio contro tutti e per sterminare tutti gli empi: e per punire le carni per le empietà da loro empiamente commesse e per tutte le dure parole dette contro di essi dagli empi peccatori. » « Profetò pure di essi Enoc, il settimo patriarca dopo Adamo, dicendo: "Ecco venire il Signore con le sue sante miriadi, a tenere giudizio contro tutti e punire tutti gli empi per tutte le opere di empietà da loro empiamente commesse e per tutte le dure parole dette contro di Lui dagli empi peccatori". »

I punti di vista del libro di Enoc differiscono per molti aspetti sia da quelli delle fonti rabbiniche sia da quelli degli stereotipi cristiani del giudaismo. Tali differenze sono divenute og-

Gli affreschi di una sinagoga, risalenti al III secolo e rinvenuti a Dura Europus in Siria, sono rappresentazioni allegoriche di episodi del Vecchio Testamento. L'illustrazione della pagina a fronte è un particolare del « Ciclo di Ezechiele » appartenente alla parete nord della sinagoga: vi è raffigurato Ezechiele, in abbigliamento persiano, ritto accanto a tre cadaveri. Una Psiche in stile greco protende le mani verso la testa del cadavere in alto, mentre altre tre figure dalle stesse sembianze volano verso il basso. Si è supposto che queste immagini femminili rappresentassero i quattro venti, invocati da Ezechiele per infondere nuova vita nei cadaveri. Il particolare è tratto da una fotografia degli affreschi di Frank J. Darmstaedter del Jewish Theological Seminary of America.

getto di approfondite ricerche erudite. Tra i rotoli del Mar Morto si sono rinvenuti anche undici manoscritti frammentari del libro di Enoc. J. T. Milik, che sta interpretando i manoscritti e ne cura la pubblicazione, afferma che il piú antico è stato copiato nella prima parte del II secolo a.C. Secondo questo studioso, alcuni paragrafi di Enoc sono molto piú antichi ed egli avanza persino l'ipotesi che gli ultimi curatori delle porzioni del *Genesi* nel Vecchio Testamento, relative a Enoc, possano essersi serviti di un riassunto di un'antica fonte del libro di Enoc (*Genesi* 6: 1-4). Uno dei punti piú importanti toccati dal paragrafo di Enoc, considerato da Milik il piú antico, consiste nel fatto che il peccato è il risultato dell'unione illecita tra gli angeli (figli di Dio) e le donne. Da questa unione ebbero origine non solo la conoscenza proibita, ma anche gli spiriti demoniaci.

Il rinvenimento dei rotoli del Mar Morto contribuí a ridestare l'interesse per gli scritti ebraici e cristiani non inclusi nella Bibbia. Mentre tali scritti, apocrifi e pseudoepigrafici, erano noti solo in lingua greca o da traduzioni dal greco, gran parte del materiale contenuto nei rotoli era in lingua ebraica o aramaica. La traduzione di questi libri ebbe per gli studiosi dell'epoca conseguenze profonde. Infatti, sia quelli ebraici sia quelli cristiani avevano avuto una certa tendenza a valutare le diverse espressioni religiose giudaiche nei termini delle proprie ortodossie. La cultura ebraica ha sottovalutato gli aspetti mistici, speculativi o comunque non razionali del giudaismo. Ne discese un quadro piuttosto limitato del giudaismo rabbinico che era sorto trionfante dopo la distruzione del Tempio di Gerusalemme operata dai romani nel 70 d.C. Questo quadro trovò sostenitori anche fra gli studiosi cristiani del giudaismo rabbinico, anche se costoro, tutto sommato, rivolsero la propria attenzione piuttosto agli scritti apocrifi e pseudoepigrafici.

L'influenza greca si diffuse nei territori che si affacciavano sul Mediterraneo orientale grazie alle conquiste compiute da Alessandro Magno nel IV secolo a.C. L'ellenizzazione delle culture e delle religioni orientali continuò per tutto il periodo della dominazione romana. Prove della ricchezza espressiva del giudaismo sono state rinvenute in parecchi luoghi (*punti in nero*).

Il loro interesse è rivolto in primo luogo e innanzitutto alla ricostruzione dell'ambiente del Nuovo Testamento. In questa ricostruzione il giudaismo rabbinico trova, nel migliore dei casi, una collocazione ambigua. I libri pseudoepigrafici, tuttavia, erano considerati come una « letteratura di opposizione » all'immagine caricaturale del giudaismo farisaico-rabbinico attribuito al-l'*establishment* del periodo anteriore alla distruzione del Tempio. Cosí agli pseudoepigrafi era attribuita un'importanza capitale. Con il ritrovamento dei rotoli del Mar Morto vennero alla luce le caratteristiche del sistema di vita della setta di Qumran a cui tali rotoli sono dovuti, e fu subito chiaro il contesto in cui era avvenuta la creazione di almeno una parte della letteratura pseudoepigrafica (si veda *Gli aderenti al Nuovo Patto di Qumran*, di Shemaryahu Talmon in « Le Scienze », n. 42, febbraio 1972).

Conseguenze altrettanto importanti ebbe la scoperta della biblioteca gnostica rinvenuta nel 1945 presso Nag Hammadi, in Egitto. Gli gnostici, un gruppo di sette religiose risalenti al tardo periodo greco-romano e al primo cristianesimo, cercavano la salvezza dell'anima mediante l'occultismo. Questo movimento costituí una delle maggiori sfide portate al cristianesimo ortodosso del II e del III secolo d.C. Fino alla scoperta della biblioteca, la maggior parte delle notizie che si possedevano sullo gnosticismo provenivano dagli scrittori cristiani che si erano trovati a dover combattere l'eresia. Con la traduzione del copto e la pubblicazione dei papiri gnostici, diviene sempre piú evidente che, con ogni probabilità, gran parte dello gnosticismo è di origine ebraica. Un esempio singolare è costituito da un trattato intitolato la *Testimonianza del Vero*: in esso sono mescolati frammenti piú antichi, alcuni dei quali presentano molti punti di contatto con le interpretazioni ebree ortodosse delle Scritture, sebbene la posizione teologica sia del tutto gnostica.

Il regno di Erode (*confini in grigio*), durato tra il 40 a.C. e il 4 d.C., comprendeva la maggior parte della Palestina. Sebbene Erode fosse re dei giudei, era vassallo di Roma. I samaritani vivevano presso il Monte Gerizim, mentre un altro gruppo appartenente a una setta ebraica non ortodossa si trovava a Qumran nel deserto della Giudea.

Un frammento particolarmente importante è quello che si riferisce a un passo del *Genesi* in cui l'astuto serpente induce Eva a mangiare il frutto proibito. Gli gnostici consideravano il serpente come una figura positiva, il piú sapiente degli animali e il maestro di Eva, la quale a sua volta aveva insegnato ad Adamo la vera saggezza. In aramaico, i vocaboli che designano « Eva », « rivelare », « serpente » e « animale » sono molto simili tra loro e rendono possibile un gioco di parole che lega insieme questi concetti. La versione gnostica del passo è scarsamente approfondita e il passo stesso non presenta alcun elemento cristiano; ciò dimostra che esso probabilmente è tratto da fonti giudaiche gnostiche precristiane. L'interpretazione in chiave positiva del serpente e della conoscenza ottenuta mangiando il frutto proibito è una tipica deformazione gnostica della versione ortodossa.

In questi manoscritti gnostici è possibile individuare tra gli ebrei una piú grande varietà di espressioni e manifestazioni religiose di quanto la maggior parte degli storici non sospettasse. Per esempio il giudaismo rabbinico aveva una visione dell'uomo che comportava l'esistenza del libero arbitrio e della piena responsabilità morale dell'indivi-

duo. I membri della setta ai quali sono dovuti i rotoli del Mar Morto sostenevano invece la posizione opposta, vale a dire che la natura dell'uomo e il suo destino sono predeterminati. Questa posizione è ben illustrata da un passo del Manuale di disciplina rinvenuto a Qumran: « Dal Dio che possiede la conoscenza discende ogni cosa che è e sarà, e prima che esse esistessero egli predispose per tutte un piano completo, e quando esse esistono adempiono il loro compito secondo il Suo piano glorioso che non può essere mutato. » Queste convinzioni diametralmente opposte non possono che venir considerate indicative di profonde differenze nel giudaismo dell'epoca.

Un'interpretazione del peccato originale, in seguito al quale era necessaria una qualche redenzione soprannaturale, si ritrova ne *Il quarto libro di Esdra*, un'opera ebraica scritta dopo la distruzione di Gerusalemme e del Tempio per opera dei romani. Il suo autore geme angosciato: « O tu, Adamo, cosa hai fatto? Perché, anche se sei tu che hai peccato, la mancanza non fu solo tua ma anche nostra che siamo i tuoi discendenti.» Paolo trova soluzione a un dilemma del genere nella cristologia, l'autore de *Il quarto libro di Esdra* nel riscatto promesso ai giusti nell'*escaton*. Dal punto di vista strutturale le soluzioni sono molto simili. Gli gnostici assunsero una posizione anche piú estrema: per essi il distacco dell'uomo dal mondo è assoluto; il mondo e il suo creatore sono demoniaci. Scopo dell'uomo è di liberare un minuscolo frammento di divino che è stato imprigionato in questo mondo e di provocare la sua riunione con Dio che è totalmente diverso, distaccato dal mondo stesso e inconoscibile.

Il lavoro pionieristico compiuto da Gershom G. Scholem nel delineare la storia del misticismo ebraico ha rivelato che la tradizione mistica era attiva anche nell'ambito del giudaismo rabbinico. Tuttavia nel Talmud e in altri scritti rabbinici questa tradizione trovava chiara espressione solo raramente. Scholem ha messo in evidenza le linee che, attraverso accenni e allusioni presenti nelle fonti rabbiniche, collegano le piú antiche scritture mistiche giudaiche, i libri mistici del « Carro » risalenti alla metà del primo millennio dopo Cristo, ad alcuni testi pseudoepigrafici e ai rotoli del Mar Morto. Studi recenti confermano la somiglianza tipologica che si può osservare tra i testi del « Carro » e gli scritti gnostici; inoltre vi sono tracce di influenze e stralci integrali.

Il lavoro di Scholem chiarisce perciò in modo inequivocabile la continuità della tradizione mistica in seno al giudaismo a partire dal periodo del Secondo Tempio (quello costruito da Erode nel 20 a.C. e distrutto nel 70 d.C.). Scholem ha anche dimostrato che questa tradizione non era in contrasto con il giudaismo rabbinico. Al contrario, risulta che alcuni tra i rabbini piú importanti praticavano il misticismo. Nel corso delle loro speculazioni essi adottavano una terminologia che si può far risalire fino ai rotoli del Mar Morto. È un altro fattore di cui bisogna tener conto nel valutare i fenomeni religiosi di questo periodo.

In seguito al contatto con la filosofia greca, molte culture orientali ellenizzate tentarono di esprimere le loro credenze sotto forma di sistemi ordinati e coerenti. Il metodo allegorico usato dai filosofi stoici fu applicato alle scritture bibliche, e le storie dei patriarchi si trasformarono cosí in allegorie che indicavano la via giusta da seguire. Dove la cultura ebraica entrò in contatto con quella greca, fiorirono anche l'apologetica (spesso molto violenta) e la storiografia. La vasta comunità ebrea di Alessandria si basava sulle storie di Giuseppe e di Mosè per giustificare la propria posizione in Egitto. Anzi, scrittori ebrei ellenizzati, come Artapano, attribuivano ad Abramo, Giuseppe e Mosè non solo l'invenzione della scrittura, dell'astronomia e del sistema amministrativo che governava l'Egitto, ma anche la fondazione dei culti degli animali, tipicamente egiziani. Questi culti erano aborriti sia dagli ebrei sia dai greci.

Il tempio di Araq el-Emir fu costruito da Ircano, appartenente alla famiglia dei Tobiadi, agli inizi del II secolo a.C. La ricostruzione è basata sui reperti degli scavi condotti nel sito sotto il patrocinio delle American Schools of Oriental Research. L'esistenza di questo e di altri templi ebraici indicherebbe che le diverse forme espressive del giudaismo erano tollerate,

Questo simbolo magico, a cui è associato il nome di Salomone, appartiene a un libro di magia dal titolo *Il testamento di Salomone*. L'invocazione del nome di Salomone trae origine dai riti magici ebraici, nei quali era impiegata per esorcizzare i demoni.

Dal tempo della riforma di re Giosia, se non prima, a Gerusalemme (621 a.C.), era opinione comune che i sacrifici al Dio di Israele si dovessero compiere solo nel Tempio della città santa. Secondo le fonti bibliche, si cercò di abolire altri templi sparsi per tutta la Palestina. Anche la tradizione rabbinica considera il tempio di Gerusalemme come l'unico luogo adatto al culto sacrificale ebraico. Durante tutto il periodo del Secondo Tempio vi sono tuttavia prove dell'esistenza di templi e di sacrifici compiuti fuori dalle mura di Gerusalemme e persino fuori dalla Palestina. Papiri che risalgono al V secolo a.C., rinvenuti a Elefantina, nell'Egitto meridionale, parlano di una colonia militare ebraica che possedeva un tempio proprio. I papiri riferiscono inoltre che il tempio venne distrutto nel 408 a.C. durante un massacro, e riportano una copia della petizione inviata dagli ebrei di Elefantina al governatore persiano della Giudea per chiedere il suo aiuto per la ricostruzione del tempio. La petizione afferma che una lettera dello stesso tenore era stata inviata al governatore di Samaria, dove a quel tempo esisteva un altro tempio dedicato al Dio di Israele. Il linguaggio in cui è stilata la petizione rivela che gli autori consideravano la loro richiesta, e quindi l'esistenza del tempio, del tutto naturale e legittima. Eppure sembra che il culto del tempio di Elefantina fosse sincretico, che comprendesse cioè anche divinità diverse dal Dio di Israele.

Si è affermato che gli ebrei di Elefantina avevano un tempio fuori da Gerusalemme dove praticavano un culto sincretico perché avevano assimilato pratiche pagane durante i numerosi anni trascorsi come mercenari al seguito dell'esercito persiano. Inoltre è probabile che il loro desiderio di avere un tempio sia stato determinato dal fatto che i loro antenati venivano da una zona del regno settentrionale di Israele toccata solo in parte dalla riforma di re Giosia. È anche possibile che questa ipotesi sia fondata, quantunque non particolarmente convincente. È comunque certo che quello di Elefantina era solo uno dei numerosi templi ebraici che esistevano durante tutto il periodo del Secondo Tempio.

Circa dieci anni fa ad Araq el-Elmir, in Transgiordania, è stato portato alla luce un tempio che risale al II secolo a.C. e sembra sia stato costruito da Ircano, che apparteneva alla nobile famiglia ebrea dei Tobiadi. I Tobiadi si erano messi in particolare evidenza nel secolo precedente. Nella sua opera *Antichità giudaiche*, lo storico Giuseppe Flavio riferisce gli avvenimenti che indussero Ircano a lasciare Gerusalemme e a ritirarsi sull'altra sponda del Giordano. Giuseppe parla anche di un tempio costruito in Egitto in un periodo di poco posteriore. Il sommo sacerdote del Tempio di Gerusalemme, Onias III, fu deposto dalla carica e si recò in Egitto dove, ottenuto il permesso reale, costruí un tempio in un luogo chiamato Leontopoli. Il tempio venne usato finché non fu chiuso dai romani poco dopo la distruzione del Tempio di Gerusalemme del 70 d.C.

Le *Antichità* di Giuseppe Flavio riportano anche un altro esempio di culto ebraico praticato lontano da Gerusalemme. L'autore vi trascrive un proclama emanato dai governanti della città di Sardi in Asia Minore, che risale al I secolo a.C. Questo documento dichiara tra l'altro che il consiglio della città permette che si stabilisca un culto ebraico, evidentemente un culto sacrificale.

Le testimonianze dimostrano cosí che in realtà gli ebrei costruirono templi e compirono riti sacrificali dedicati al Dio di Israele in luoghi diversi dal Tempio di Gerusalemme. Considerando

Le catacombe di Beth Shearim, nella Palestina centrale, vennero usate dagli ebrei come luoghi sepolcrali dal II al IV secolo d.C. I lunghi corridoi in pendenza erano stati scavati nella dura roccia. Le pareti erano decorate con incisioni.

Un guerriero recante una *menorah*, il candelabro usato nei riti ebraici, fa parte delle incisioni parietali raffiguranti simboli religiosi delle catacombe di Beth Shearim. Il significato simbolico può essere ricercato in un qualche rapporto, forse di tipo religioso, del soldato con gli scopi e le speranze del giudaismo.

Torah Shrine, presso le catacombe di Beth Shearim nella valle di Jezreel, contiene numerosi simboli religiosi, tra cui una conchiglia a pettine, un leone che porta una bardatura (*in alto a destra*) e un *lulab* (*a destra, tra le colonne*), un fascio di rami di palma che veniva usato nella Festa dei Tabernacoli.

l'opinione largamente diffusa secondo cui la sola Gerusalemme aveva diritto a un tempio, diritto stabilito almeno quattro secoli prima di Ircano e Onias, il fatto che ne esistessero altri dovrebbe far riflettere. Significa forse che le fonti ebraiche che rivendicavano questo diritto, soprattutto i libri del Vecchio Testamento, sono tendenziose, che riportano come dato di fatto quanto era solo nel desiderio dei loro autori? Ciò sarebbe portare le conclusioni all'estremo. È comunque certo che le fonti bibliche non offrono un quadro completo di quello che era tollerato: nel giudaismo c'era maggior varietà di quanto esse non rivelino.

Benché sia noto che in ogni religione superiore esiste una componente di magia, l'enorme importanza di questa ultima nella religione ebraica durante il periodo ellenistico-romano è degna di rilievo. Il libro di Enoc ne rintraccia l'origine negli insegnamenti proibiti rivelati dagli angeli caduti che avevano avuto rapporti con le figlie dell'uomo. La sua posizione è comunque quella di un chiaro rifiuto delle pratiche magiche. Sembra che altri gruppi credessero fermamente nell'esistenza di demoni e di spiriti maligni, e ad altri aspetti della pratica della magia, come risulta da vari accenni in importanti scritti di questo periodo. Una tradizione magica molto diffusa riguardava re Salomone, che aveva incatenato i demoni obbligandoli a contribuire alla costruzione del Tempio: il suo nome veniva frequentemente invocato negli esorcismi contro i demoni ritenuti responsabili dei mali degli uomini, e si ritrova anche in simboli magici e amuleti del periodo.

In epoca recente è stato scoperto un manoscritto medioevale particolarmente significativo, il *Sefer Harazim* (*Il libro dei misteri*). Questo libro, un'opera di magia scritta in ebraico e risalente al III secolo, fornisce una connessione tra la magia pagana e quella ebraica. Tipica della mescolanza di elementi pagani ed ebraici presente nel libro è la preghiera a Helios, il dio del Sole greco, che è composta in lingua greca, ma scritta in caratteri ebraici. Ovviamente i gruppi che produssero il libro ritenevano di poter accogliere elementi pagani nella propria dottrina e nella pratica religiosa.

Sarebbe opportuno commentare i dipinti ebraici e le miniature dei manoscritti del tardo periodo ellenico e romano. Tra il I secolo a.C. e il I secolo d.C. non esiste alcuna forma d'arte figurativa ebraica proveniente da Gerusalemme o da altre località ebraiche della Palestina. Il motivo sembra vada ricercato nell'interpretazione del secondo comandamento: « Non ti devi fare immagini scolpite ». Questa interpretazione non era l'unica. Essa è in contrasto con la situazione dei secoli successivi, ai quali si fa risalire una serie di pavimenti di sinagoga portati alla luce in tutta la terra di Israele. Questi pavimenti sono decorati con varie immagini, che in alcuni casi includono anche quelle di divinità pagane. Molti sarcofagi del cimitero ebraico di Beth Shearim, nella valle di Jezreel, sono abbelliti da opere di arte figurativa.

Inoltre, ricerche recenti compiute su manoscritti illustrati del Vecchio Testamento, di epoca cristiana, e i notevoli affreschi in una sinagoga del III secolo a Dura Europus in Mesopotamia hanno dimostrato che esistevano manoscritti biblici ebraici in lingua greca forniti di illustrazioni che probabilmente risalivano a un'epoca precristiana. La tradizione iconografica e illustrativa dei manoscritti biblici ebraici venne ripresa dall'arte cristiana nel corso del suo sviluppo, ed entrò cosí nelle tradizioni culturali di Bisanzio e dell'Europa. I manoscritti ebraici derivavano tecniche e motivi da illustrazioni greche contemporanee di opere classiche. Questa tradizione di per se stessa riflette i vari atteggiamenti nei confronti del secondo comandamento ed è degna di interesse. Inoltre sembra porre interrogativi piú vasti sulle antiche somiglianze e differenze esistenti tra gli ebrei della Palestina e della Diaspora e sulle relazioni tra la cultura ellenistica ebraica e la cultura ellenistica pagana.

In ultima analisi, è probabile che il punto cruciale sia quello dell'« equilibrio di potere », delle relazioni tra queste diverse forme (o anche categorie) di espressione religiosa in seno all'edificio del giudaismo. Ne consegue che il problema centrale sarà costituito dalla descrizione globale del giudaismo stesso. Per gli storici del cristianesimo, una migliore comprensione del giudaismo contemporaneo chiarirà non solo il contesto della predicazione di Gesú e dell'attività dei suoi discepoli, ma anche quello della diffusione della nuova fede in un mondo piú vasto. Per lo storico del giudaismo, o della religione in generale, un riesame piú meditato chiarirà come l'espressione religiosa si sviluppasse in modo barocco e disordinato e quanto fosse sperimentale e creativa per tutto questo periodo. Durante il suo lavoro, lo storico si troverà di fronte, immensamente varia e feconda, la religione ebraica di un'età che fu cosí importante per tutte le generazioni successive.

Marte visto da Mariner 9

I dati forniti per circa un anno da Mariner 9 *paiono indicare che l'interno di Marte sta ora iniziando a riscaldarsi; le frequenti incisioni costituiscono il risultato dell'erosione da parte dell'acqua o di qualche altro agente*

di Bruce C. Murray

Poco piú di un anno fa, nel novembre 1971, una complessa sonda automatica, il *Mariner 9*, accese i suoi razzi frenanti ed entrò in orbita attorno a Marte: divenne in tal modo il primo oggetto costruito dall'uomo, in orbita attorno a un pianeta diverso dalla Terra. Continuando a percorrere la sua orbita, che lo portava a una distanza variabile tra un minimo di 1650 e un massimo di 17 100 chilometri dalla superficie di Marte, *Mariner 9* cominciò a inviare a terra immagini e informazioni scientifiche con un ritmo rimasto pressoché stazionario per circa un anno. In questo periodo *Mariner 9* ha fornito una quantità di dati circa 100 volte superiore a quella accumulata da tutte le precedenti missioni verso Marte. Di conseguenza, è anche mutata in modo drastico l'immagine che generazioni di astronomi e di scrittori si erano fatta del pianeta che per lo piú era ritenuto simile alla Terra. In seguito alla missione di *Mariner 9* è ora possibile formulare, sulla costituzione geologica di Marte, ipotesi dello stesso tipo di quelle che venivano formulate a proposito della Luna all'inizio degli anni '60.

Sarà bene ricordare che le prime immagini ravvicinate di Marte, riprese nel 1965 da *Mariner 4*, mostrarono una superficie la cui struttura morfologica principale era costituita da grandi crateri che ricordavano la squallida superficie lunare. Quattro anni piú tardi, le immagini restituite da *Mariner 6* e *7* mostrarono che la superficie di Marte non era uniformemente cosparsa di crateri: vi comparivano invece anche ampie zone a morfologia caotica, che non avevano analogo né sulla Terra né sulla Luna. Inoltre, apparve chiaro che la vasta area tondeggiante, da lungo tempo nota agli astronomi terrestri come il « deserto » Hellas, era, almeno nei limiti del potere risolvente delle telecamere dei *Mariner*, pressoché priva di qualunque struttura superficiale. Nessuna delle immagini restituite dai *Mariner* citati mostrò nulla che indicasse l'esistenza di una certa attività vulcanica: ciò faceva supporre che Marte fosse inattivo dal punto di vista tettonico.

Questa concezione ha dovuto essere radicalmente modificata sulla base delle immagini restituite da *Mariner 9*. Il nuovo panorama del pianeta si è rivelato lentamente, mano a mano che le nubi di polvere, che per settimane ne oscurarono la superficie, andavano depositandosi. Fra le altre cose apparvero nella loro imponenza quattro edifici vulcanici che sono piú grandi di qualunque struttura analoga conosciuta sulla Terra. Le immagini di *Mariner 9*, inoltre, mostrano un ampio sistema di valli principali e tributarie e numerose incisioni sinuose che, a prima vista, si presentano come se fossero state scavate da acqua corrente. In nessun altro luogo della superficie del pianeta vi è altra indicazione di erosione da parte dell'acqua. È questo, probabilmente, il mistero principale che rimane dopo il pieno successo della missione di *Mariner 9*.

Progettato e costruito dal Jet Propulsion Laboratory dell'Istituto di Tecnologia della California, che ha curato anche le sonde precedenti, *Mariner 9* fu colmato di strumenti e congegni elettronici. Dopo aver bruciato oltre 400 chilogrammi di carburante, che gli hanno consentito di percorrere quasi 460 milioni di chilometri in 167 giorni, *Mariner 9* pesava poco piú di 600 chilogrammi quando finalmente entrò in orbita attorno a Marte. Le telecamere e le apparecchiature di bordo sono state progettate da diversi gruppi di ricerca appartenenti a laboratori governativi e a piú di una ventina di università statunitensi. Il gruppo addetto alle riprese televisive, al quale ho appartenuto, era guidato da H. Masursky dell'US Geological Survey ed era costituito da circa 30 membri. Gruppi appena piú piccoli avevano la responsabilità di progettare e di analizzare i dati relativi allo spettrometro ultravioletto, al radiometro infrarosso e allo spettrometro interferometrico infrarosso. Altri gruppi avevano lo scopo di analizzare i dati relativi alla traiettoria (che dovevano fornire informazioni sulle anomalie gravitazionali di Marte) e quelli forniti dalle circa 100 occultazioni dei segnali radio della sonda (che dovevano consentire di approfondire la conoscenza dell'atmosfera e della superficie del pianeta).

Marte prima di Mariner 9

La convinzione tradizionale secondo cui Marte doveva essere simile alla Terra, non era priva di giustificazioni obiettive. Marte compie una rotazione completa attorno al proprio asse in 24,5 ore e il suo asse ha un'inclinazione sul piano dell'eclittica quasi esattamente identica a quella della Terra; ciò costituisce una base astronomica comune per ciò che riguarda i mutamenti stagionali nella quantità di radiazione solare ricevuta dai due emisferi del pianeta. Marte ha due calotte polari bianche, che originariamente si riteneva fossero costituite d'acqua, le quali si alternano da un emisfero all'altro una volta ogni anno marziano (687 giorni terrestri). Il pianeta mostra anche strutture scure e chiare che mutano di aspetto in stretta relazione con le stagioni.

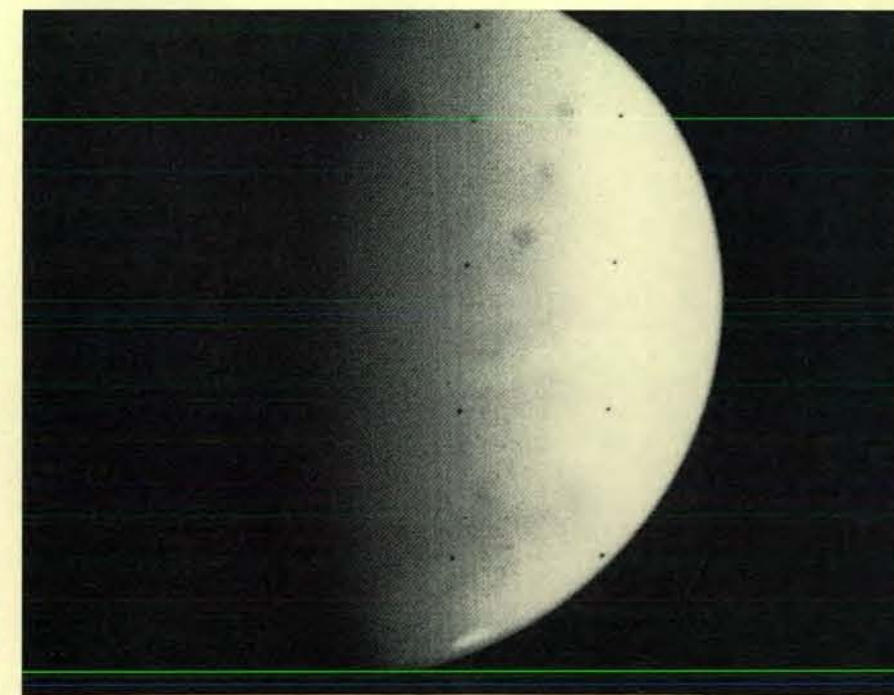
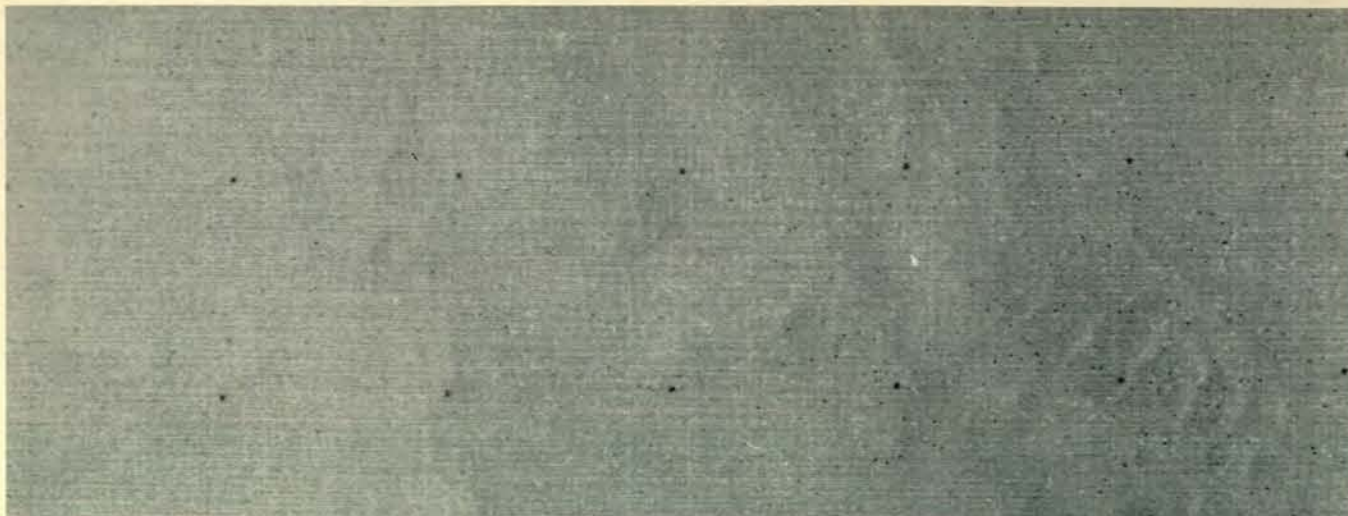
I primi astronomi che si occuparono del problema avevano pensato che le strutture scure potevano essere costituite da vegetazione. Scienziati successivi e piú cauti ritennero plausibile che Marte avesse avuto una storia primordiale simile a quella della Terra; ciò implicava l'esistenza di oceani e di un'atmosfera con un contenuto d'acqua sufficiente a determinare precipitazioni che avrebbero esercitato una azione erosiva sulla superficie. Si è poi supposto che, a causa della piccola massa di Marte (un decimo della massa della Terra) e della bassa gravità, tale atmosfera con il suo contenuto d'acqua si sia poi allontanata dal pianeta lasciandolo nello stato di aridità che oggi lo caratterizza. Tale concezione, che faceva di Marte un pianeta assai simile alla Terra, influenzò notevolmente le proposte che, all'inizio dello sviluppo della ricerca spaziale, furono avanzate nella direzione di un'esplorazione biologica del pianeta. Sembrava ragionevole supporre che la vita potesse aver avuto origine su Marte cosí come era accaduto sulla Terra, in conseguenza dell'elevata concentrazione che le molecole organiche abiogeniche potevano avere raggiunto negli oceani primitivi. Una volta che la vita fosse sbocciata sul pianeta, almeno i microorganismi avrebbero ben potuto essere in grado di adattarsi ai mutamenti delle condizioni ambientali; cosí essi avrebbero potuto anche sopravvivere fino a che le sonde automatiche lanciate da Terra non li avessero scoperti e analizzati.

Tali aspettative furono deluse da quanto rivelò *Mariner 4*: non solo il panorama marziano appariva squallido e simile a quello lunare ma si scoprí anche che il pianeta mancava di un campo magnetico che avrebbe potuto fare da scudo contro le particelle che il Sole diffonde nello spazio. Per di piú la pressione atmosferica di Marte risultò essere inferiore all'1 % di quella terrestre: un valore almeno 10 volte minore di ogni stima precedente. Dato che la forza di gravità alla superficie di Marte è piú di un terzo di quella alla superficie terrestre, Marte avrebbe potuto facilmente trattenere un'atmosfera la cui pressione in superficie fosse un decimo della pressione esercitata dall'atmosfera sulla superficie terrestre.

Tali osservazioni furono estese da *Mariner 6* e *Mariner 7*. Queste due sonde confermarono che le calotte polari sono costituite di biossido di carbonio solido purissimo, cioè di ghiaccio secco e non di ghiaccio d'acqua. Le immagini che rivelarono la presenza su Marte di un terreno a morfologia caotica, indicarono che alcune parti del pianeta avevano subito fenomeni di dislocazione e che esso doveva essere stato interessato da una certa attività endogena. Di conseguenza, alcuni ricercatori supposero che il pianeta potesse proprio ora essere soggetto a riscaldamento interno, come alcuni modelli termici di Marte avevano già suggerito in modo del tutto indipendente. L'opinione prevalente dei ricercatori che si occupavano dei *Mariner* era tuttavia ancora quella secondo cui Marte doveva essere sostanzialmente assai piú simile alla Luna che non alla Terra. Da allora le strutture chiare e scure che apparivano quando si osservava Marte al telescopio furono in genere attribuite a una qualche interazione dell'atmosfera con la polvere. Nel gruppo di immagini inviate a terra dai *Mariner 6* e *7* apparvero indizi che tali interazioni erano in qualche modo connesse con strutture morfologiche locali, ma non fu possibile fornirne una spiegazione piú generale. Malgrado ciò, alcuni ricercatori giunsero alla convinzione che tali strutture potessero invece dipendere da variazioni nell'umidità del suolo.

Il 1971 doveva vedere l'esplorazione di Marte attraverso due sonde, *Mariner 8* e *Mariner 9*, entrambe in orbita attorno al pianeta. Lo scopo delle due sonde era di cartografare la maggior parte della superficie del pianeta con una risoluzione sufficientemente elevata da rivelare i processi sia esogeni sia endogeni in atto, per studiare i fenomeni transitori sulla superficie e nell'atmosfera e per fornire un rilevamento che durasse un tempo sufficientemente lungo (da 9 mesi a un anno) da poter osservare i cambiamenti stagionali sulle strutture superficiali in modo da chiarirne le origini. Quando *Mariner 8* andò distrutto durante le operazioni di lancio, divenne indispensabile modificare gli obiettivi di *Mariner 9* in modo che questo potesse svolgere il lavoro previsto per entrambe le sonde.

Mariner 9 raggiunse Marte il 13 novembre 1971; in quel momento la piú grande tempesta di polvere osservata in piú di un secolo stava infuriando sul pianeta oscurandone quasi totalmente la superficie. Le prime immagini, riprese da una distanza di diverse centinaia di migliaia di chilometri, non mostrarono altro che un debole barlume della calotta polare meridionale (*si veda l'illustrazione in basso*).

La tempesta di polvere interessò soprattutto i ricercatori che miravano a studiare l'atmosfera del pianeta (c'era infatti da sperare che questo fenomeno rivelasse in che modo le particelle di polvere potessero essere tenute in sospensione da un mezzo cosí rarefatto) ma costituí una delusio-

Un'immagine di Marte, ripresa da *Mariner 9* alla distanza di 400 000 chilometri, un giorno e mezzo prima di raggiungere l'orbita marziana, il 13 novembre 1971: il pianeta è oscurato da una coltre di polvere. La piú importante tempesta di polvere osservata in piú di un secolo aveva infatti nascosto tutte le strutture superficiali salvo la calotta polare meridionale, che stava rapidamente riducendosi con l'avvento della primavera nell'emisfero meridionale, e le quattro macchie scure visibili nel quadrante superiore destro; queste ultime sono la Nix Olympica e altri tre edifici vulcanici.

I risultati dell'elaborazione al calcolatore sono dimostrati da questa serie di copie della stessa immagine, ripresa presso il polo sud marziano. L'elaborazione eseguita nello spazio di 5 minuti dal Mission Test Computer del Jet Propulsion Laboratory, è stata applicata a tutte le immagini ricevute da *Mariner 9*. Ciascuna di esse veniva trasmessa secondo segnali radio codificati, nei quali la luminosità di ciascun punto era rappresentata da una sequenza di 9 numeri binari piuttosto che non dal livello di intensità (segnale AM) o dal tono di frequenza (segnale FM). Ciascuna immagine era costituita da 700 linee, a loro volta fatte di 832 elementi per linea. Il segnale ricevuto veniva registrato su un nastro magnetico e simultaneamente immesso in un tubo catodico. Nella forma iniziale (*in alto*) l'immagine appare abbastanza grigia e priva di strutture, poiché il reale contrasto sulla superficie di Marte è abbastanza basso. Il primo stadio dell'elaborazione al calcolatore conduce a una versione corretta per le ombre (*al centro*), che indica la luminosità relativa, dandole un qualche risalto. La porzione destra dell'immagine è scura poiché è al di sotto del terminatore e immersa nell'ombra. Un secondo livello di elaborazione (*in basso*) sopprime le differenze a grande scala nella luminosità allo scopo di esaltare soprattutto i dettagli topografici, sia nelle zone luminose sia in quelle scure. L'immagine ora mostra una serie di complesse strutture topografiche in cui diverse dorsali distanti l'una dall'altra da due a cinque chilometri, si incontrano; è la struttura che il gruppo di ricerca *Mariner* ha soprannominato Città Inca. Le dorsali sembrano essere costituite di materiale resistente che ha riempito una serie di fratture e che poi è stato sottoposto all'erosione, resistendo più del terreno circostante. Ripresa effettuata da *Mariner 9* a 2937 chilometri di distanza.

ne per coloro che si interessavano alle strutture superficiali del pianeta. Per esempio, era stato previsto di riprendere una serie di immagini in sequenza, che avrebbe poi dovuto essere stampata a colori, che mostrava il pianeta che diveniva via via più grande; questa sequenza avrebbe costituito una sorta di ponte visivo tra il livello di dettaglio visibile attraverso i telescopi terrestri e quello che poi sarebbe stato visibile quando la sonda fosse stata in orbita attorno a Marte. Per quanto riguarda le immagini di *Mariner 7*, in effetti, utilizzando una serie di filtri, rosso, verde e blu, è stato effettuato un tentativo di realizzare una sequenza di immagini di questo tipo atta a mostrare Marte nei suoi colori naturali.

I grandi vulcani

La tempesta di polvere ritardò di circa 3 mesi il lavoro di cartografia sistematica della superficie marziana. Anche durante la tempesta, tuttavia, nelle prime immagini riprese dalla sonda in orbita furono viste ripetutamente quattro chiazze scure nella zona equatoriale del pianeta. Esse chiaramente rappresentavano strutture superficiali permanenti di altezza sufficiente a farle emergere dalla coltre di polvere. Probabilmente apparivano scure semplicemente perché la loro riflettività superficiale era inferiore a quella dell'atmosfera allora ricca di polvere e perciò assai riflettente.

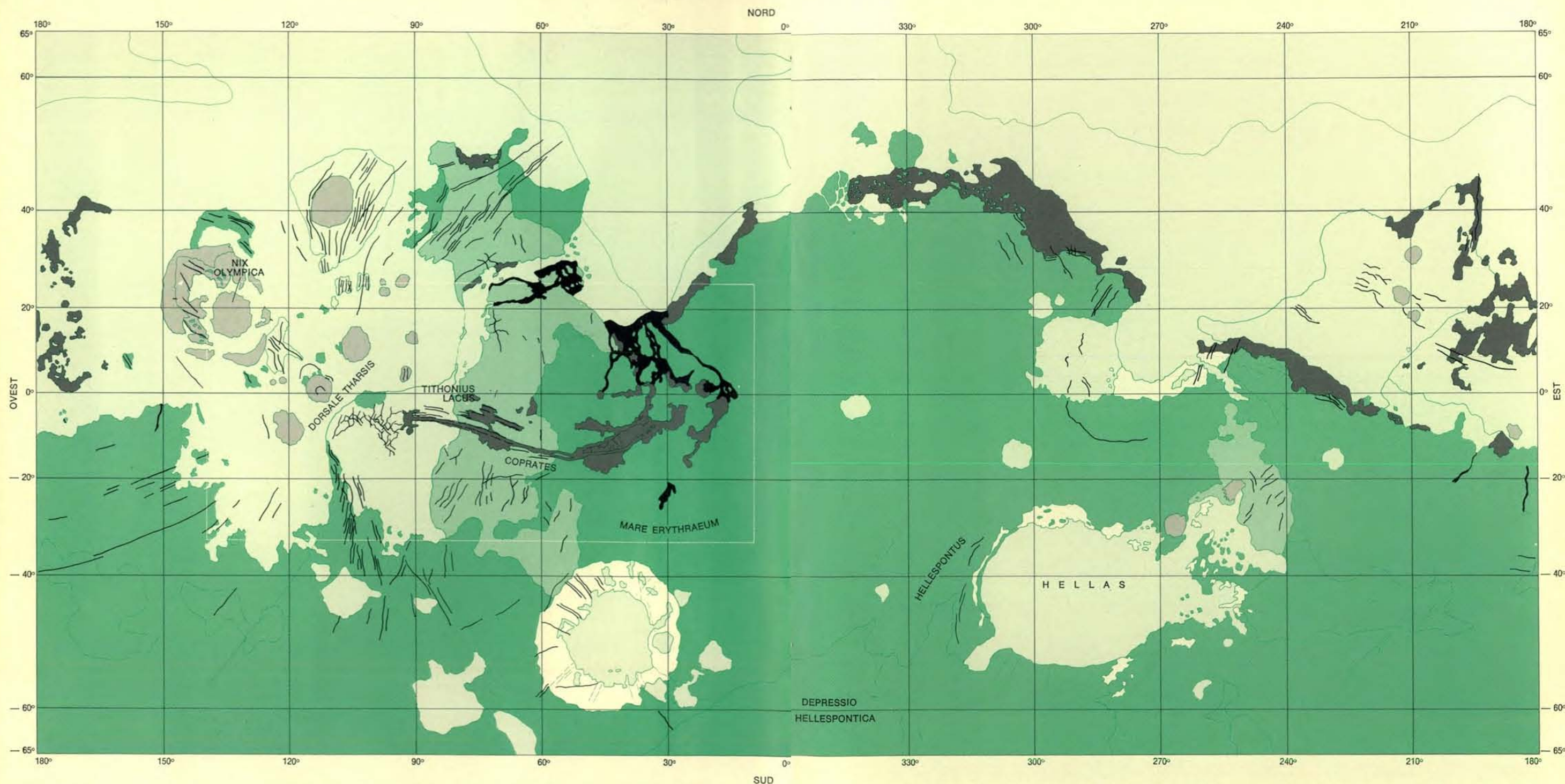
Una delle quattro chiazze corrispondeva alla posizione in cui doveva trovarsi la Nix Olympica, così denominata perché era una struttura normalmente visibile da terra e caratterizzata da elevata riflettività e variabilità nel tempo. Quando questa chiazza fu osservata con la telecamera ad alta risoluzione, apparve un'immagine sorprendente: ciò che si vide fu la caratteristica immagine di una caldera vulcanica. Le caldere non sono strutture rare sulla Terra; ne sono un esempio alcune delle isole hawaiiane. La caldera marziana, tuttavia, era circa 30 volte più grande, come diametro, di qualunque altra presente nelle Hawaii. Quando la tempesta di polvere si placò, fu possibile osservare per intero la Nix Olympica: un edificio vulcanico enorme, la cui base aveva un diametro di oltre 500 chilometri ed era cioè assai più grande di qualunque struttura analoga nota sulla Terra (*si veda l'illustrazione a pagina 66*). Le rilevazioni della pressione atmosferica eseguite successivamente con l'aiuto dello spettrometro ultravioletto e di altre tecniche, hanno dimostrato che la Nix Olympica ha un'altezza di almeno 15 chilometri e forse di 30. A titolo di paragone, si tenga conto che il Mauna Loa, il più grande edificio vulcanico delle Hawaii, non raggiunge l'altezza di 10 chilometri dal fondo del Pacifico sul quale si eleva. Le immagini ad alta risoluzione rivelarono che le altre tre chiazze scure erano anch'esse costituite da strutture vulcaniche, un poco più piccole della Nix

L'entrata in orbita di *Mariner 9* è avvenuta il 13 novembre 1971 dopo che i retrorazzi della sonda avevano generato una controspinta di circa 136 chilogrammi per 15 minuti e 15,6 secondi; il viaggio per superare la distanza di quasi 460 milioni di chilometri è durato 167 giorni. *Mariner 9* è entrato in orbita marziana da sotto il piano equatoriale con un'inclinazione di 64 gradi rispetto a quest'ultimo. L'orbita iniziale passava a una altezza variabile tra un minimo di 1385 chilometri e un massimo di 17 300 sulla superficie del pianeta. Una correzione successiva ha ridotto l'altezza massima a 17 100 chilometri portando invece quella minima a 1650 chilometri; questa operazione ha così consentito di ottenere il periodo orbitale desiderato che era esattamente di 11,98 ore.

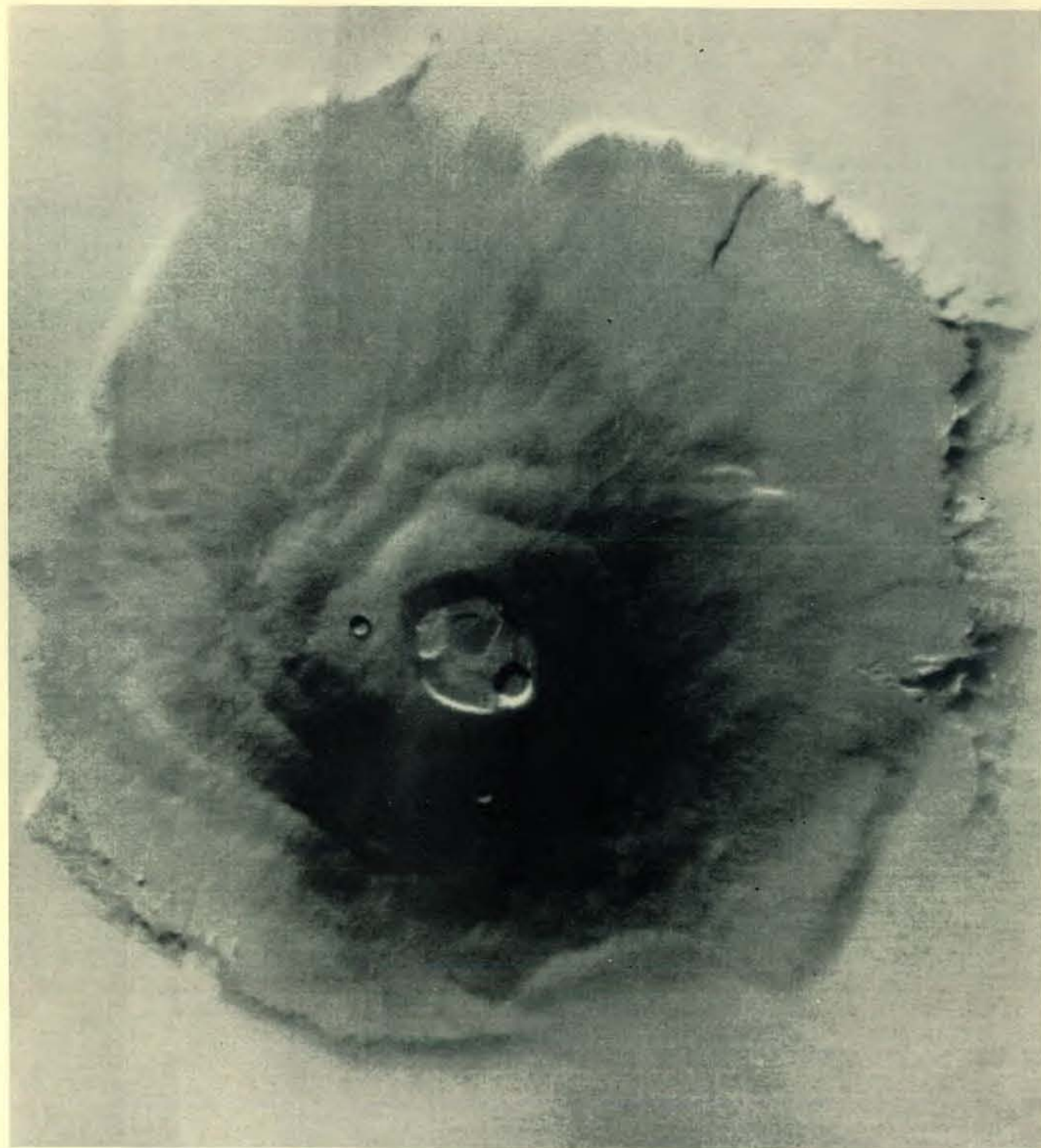
Il satellite più interno di Marte, Phobos, fotografato da *Mariner 9* da una distanza di 5500 chilometri; Phobos ha una lunghezza massima di circa 25 chilometri e una larghezza di 20 chilometri. Il satellite esterno Deimos è grande circa la metà di Phobos. L'immagine di Phobos è stata notevolmente migliorata dall'elaborazione al calcolatore.

Olympica, ma allineate fra loro in modo da formare una lunga dorsale vulcanica. Seguendo la nomenclatura tradizionale, quella struttura è ora chiamata dorsale Tharsis (*si veda l'illustrazione in queste due pagine*).

La prima struttura riconoscibile ripresa da *Mariner 9* poneva una questione affascinante: si trattava in pratica di spiegare come mai, mentre un intero emisfero del pianeta, quello che era stato osservato dai tre *Mariner* precedenti, non mostrava quasi alcuna traccia di attività endogena, la prima area ripresa in dettaglio sulla faccia opposta del pianeta mostrava invece quattro enormi edifici vulcanici. La spiegazione sembrerebbe essere che Marte sta solo ora cominciando a riscaldarsi all'interno e quindi a sviluppare la sua attività vulcanica. Probabilmente questo fenomeno è ora abbastanza avanzato nella zona Nix Olympica-Tharsis ma non si è ancora diffuso al pianeta nel suo insieme. Può darsi che in questo momento noi siamo testimoni di quella fase, ora in atto su Marte, che la Terra ha probabilmente attraversato ai primordi della sua storia, una fase le cui tracce sono state interamente cancellate sul nostro pianeta dai successivi processi eruttivi e sedimentari.

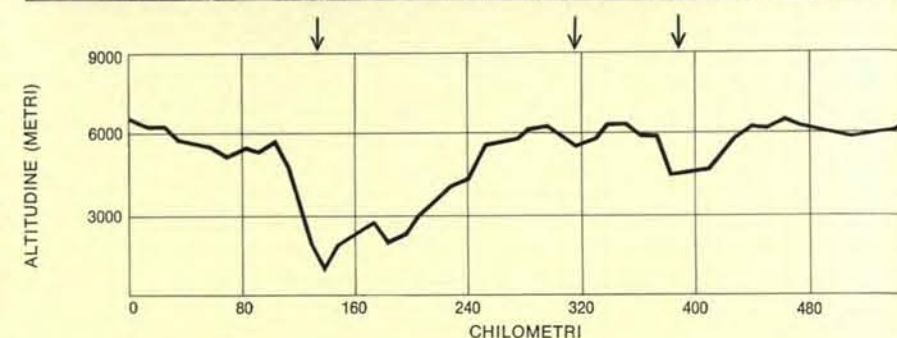
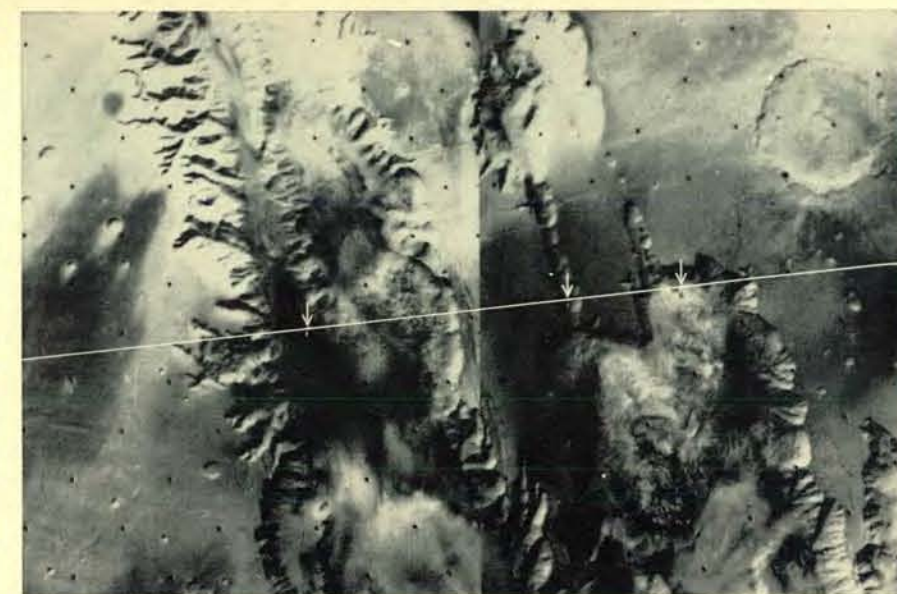
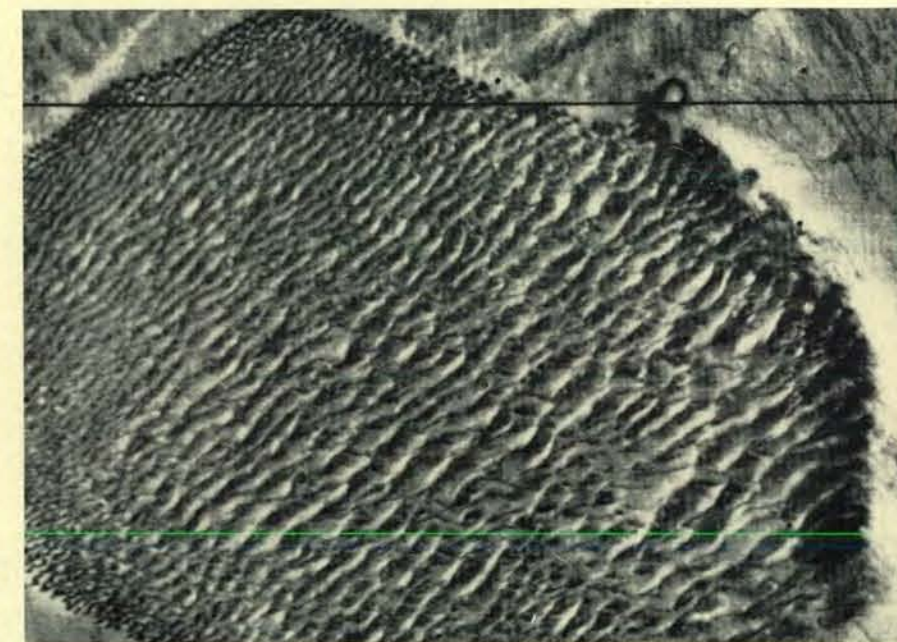
La velocità alla quale l'interno di un pianeta va riscaldandosi dipende da un certo numero di fattori ma soprattutto dalla quantità di materiale radioattivo presente nel momento in cui è avvenuta l'accrezione e dalla massa totale del corpo celeste, che determina la pressione interna e il grado di isolamento. In termini generali, se Marte avesse avuto la stessa composizione originaria che aveva la Terra, ci sarebbe da aspettarsi che il suo riscaldamento si manifesti piú lentamente poiché quel pianeta possiede soltanto un decimo della massa totale della Terra. La dimensione assoluta della Nix Olympica indica la possibilità che nell'interno del pianeta siano attive correnti di convezione profonde, cioè quegli stessi fenomeni che di qui a qualche centinaio di milioni di anni potrebbero condurre a un meccanismo dinamico del tipo della tet-

La carta geologica di Marte, uno sforzo per classificare le strutture che costituiscono la superficie del pianeta. Il rettangolo bianco racchiude la regione rappresentata nel mosaico di fotografie riportato alle pagine 70 e 71. Le zone in colore tenue sono pianure uniformi; le zone in colore di media intensità sono aree a crateri; le zone in colore intenso sono regioni antiche a crateri; le aree bianche sono costituite da terreno montuoso; quelle in grigio chiaro sono zone vulcaniche e comprendono la Nix Olympica e i tre vulcani che formano la dorsale di Tharsis. Le zone in grigio scuro indicano un terreno che è stato in qualche modo modificato, mentre quelle in nero sono depositi connessi con le incisioni. Le linee brevi interrotte sono supposte faglie. In questa carta sono classificate diverse strutture descritte nel testo o illustrate nelle immagini che accompagnano questo articolo. Per esempio, nella fotografia a colori a pag. 69, è riportata una porzione della Depressio Hellespontica; un canyon di Coprates appare a pag. 77, uno di Tithonius Lacus appare in alto a pag. 67, mentre in basso nella stessa pagina appare un cratere di Hellespontus. A pag. 79 è riportata una valle sinuosa nel Mare Erythraeum. Hellas è una zona pressoché priva di strutture con un diametro di circa 1600 chilometri. Questa carta è stata preparata da M. Carr, J. F. McCauley, D. Milton e D. Wilhelms dell'US Geological Survey insieme con il gruppo di ricerca dell'esperimento televisivo del *Mariner 9*.

tonica globale che sulla Terra determina la lenta ma continua deriva dei continenti.

Immediatamente a oriente della provincia vulcanica vi è una zona intensamente fratturata e al di là di questa vi è un'altra straordinaria struttura morfologica: una serie di profondi canyon che si estendono da est a ovest lungo l'equatore (*si veda l'illustrazione alle pagine 70 e 71*). Questi canyon hanno un'ampiezza compresa tra 80 e 120 chilometri e una profondità da 5 a 6,5 chilometri e sono molto piú grandi di ogni altra struttura analoga presente sulla Terra. Anche in questo caso è necessario presupporre che la loro origine sia legata a un'attività endogena abbastanza recente. Presumibilmente, fenomeni tettonici su grande scala, orientati in senso est-ovest, hanno esposto gli strati subsuperficiali del pianeta, la cui composizione può avere in qualche modo scatenato un qualunque processo erosivo.

La Nix Olympica in un mosaico di immagini riprese da *Mariner 9* opportunamente elaborate. L'immagine è stampata con il nord sulla destra cosicché la luce del Sole sembra provenire dall'alto; in questo modo si percepisce immediatamente la struttura come un cono rilevato piuttosto che non come una depressione. L'edificio vulcanico, che si solleva al di sopra di una grande pianura relativamente uniforme, ha un diametro di base di 500 chilometri, ed è quindi assai piú grande di qualunque montagna vulcanica analoga sulla Terra. I dati relativi alla pressione indicano che la Nix Olympica ha un'altezza compresa tra 15 e 30 chilometri; non tutti sono d'accordo sul suo valore preciso. La grande caldera sommitale ha un diametro di 65 chilometri.

Un'ipotesi è che in questi fenomeni sia coinvolto anche qualche strato profondo permanentemente gelato, forse associato con l'arrivo presso la superficie di acque giovanili che preannunciavano e accompagnavano la risalita verso la superficie del pianeta del magma coinvolto nell'episodio vulcanico evidente piú a ovest. Marte è ovunque al di sotto del punto di congelamento almeno fino a una piccola profondità al di sotto della superficie. Una volta che il suolo permanentemente gelato (permafrost) fosse stato esposto all'atmosfera, il suo contenuto in acqua sarebbe sublimato, determinando la formazione di un materiale friabile, incoerente e sufficientemente mobile da servire come agente erosivo in un processo di trasporto di massa. A questo punto è necessario chiedersi dove questo materiale sia finito. Una possibilità è che i venti di Marte l'abbiano trasportato sotto forma di polvere in altre località del pianeta; è opportuno infatti tener presente che, anche se l'atmosfera marziana è rarefatta, i suoi venti possono spostarsi a una velocità di centinaia di chilometri l'ora. L'ipotesi alternativa è che quel materiale possa ancora essere scoperto in qualche luogo a occidente dei canyon. Una terza possibilità ancora è che quel materiale possa anche essere scomparso all'interno del pianeta nel quadro di un complesso meccanismo di scambi.

Il piú grande dei canyon corrisponde a una struttura da lungo tempo conosciuta col nome di Coprates, il cui aspetto muta talora con le stagioni. Osservando questo canyon nel momento in cui terminava la tempesta di polvere, si è potuto apprendere qualcosa di nuovo a proposito della variabilità del suo aspetto. Il canyon infatti è cosí profondo che uno strato considerevole di polvere rimane sospeso nell'atmosfera tra le due pareti anche dopo che l'atmosfera nelle regioni adiacenti è tornata ad essere relativamente trasparente (*si vedano le illustrazioni a pagina 77*). La polvere in sospensione nell'atmosfera rende il canyon assai piú brillante nell'aspetto che non le regioni adiacenti. Non appena però l'atmosfera nel canyon si libera della polvere, il contrasto tra l'interno e l'esterno di questa struttura diventa minimo. Perciò le strutture « superficiali » associate a Coprates, probabilmente non hanno nulla a che fare con la superficie stessa.

Altri fenomeni atmosferici di questo

Un abisso che si apre nel Tithonius Lacus all'estremità occidentale del sistema di canyon che corre lungo l'equatore di Marte. Lo spettrometro all'ultravioletto posto su *Mariner 9* ha misurato la pressione atmosferica sulla superficie del pianeta, lungo la traccia indicata (*in alto*) dalla linea bianca. I dati relativi alla pressione sono stati poi convertiti nel profilo di profondità che compare qui sopra. Il dislivello tra il punto piú alto e il punto piú basso supera i 6000 metri: il baratro aperto nel Tithonius Lacus è quattro volte maggiore di quello del Gran Canyon dell'Arizona. Tithonius Lacus si trova immediatamente a sud dell'equatore marziano a circa 85° di longitudine ovest.

La telecamera di *Mariner 9* ha scoperto, entro un cratere del diametro di 150 chilometri nella regione di Hellespontus, questo campo di dune del diametro di circa 70 chilometri. Le creste delle dune sono alla distanza di circa 1,5 chilometri l'una dall'altra.

tipo possono spiegare anche diverse altre strutture variabili precedentemente attribuite a cambiamenti stagionali della superficie. Vi sono altri tipi di variazioni che non sono spiegabili in modo cosí semplice, ma comunque esse sono sempre frutto dell'interazione fra polvere, topografia e atmosfera.

Come i vulcani, i grandi canyon di Marte indicano un episodio abbastanza recente nella storia del pianeta, in cui sono accaduti importanti eventi. La Terra è caratterizzata dalla presenza di uno stato relativamente stazionario tra processi erosivi e processi costruttivi: perciò il panorama terrestre mostra una considerevole varietà di situazioni morfologiche, da paesaggi giovani a paesaggi maturi. Nel caso dei canyon di Marte, invece, l'erosione non sembra essere controbilanciata da corrispondenti processi costruttivi, né alla morfologia giovane dei canyon si accostano paesaggi con forme mature.

Le incisioni

L'estremità orientale dei canyon raggiunge una vasta area a morfologia caotica, una piccola porzione della quale fu intravvista da *Mariner 6*. La morfologia superficiale di quest'area indica decisamente che essa è il risultato di un qualche tipo di collasso che a sua volta è geneticamente connesso con i canyon che si sviluppano piú a occidente. Una struttura particolare si estende al di fuori del terreno a morfologia caotica: si tratta di alcuni straordinari canali che si allungano in direzione nord-ovest e che sono stati anche trovati in un certo numero di altre località sul pianeta. È difficile osservare la immagine di questi canali senza considerare la possibilità che essi siano stati scavati da acqua corrente; in effetti, alcuni miei colleghi ritengono che questa sia l'unica spiegazione ragionevole.

L'età delle incisioni si può stimare osservando il rapporto dimensione-frequenza dei crateri di impatto che interessano il loro fondo. Le incisioni sono chiaramente piú giovani del terreno a crateri che si estende sulla maggior parte del pianeta, anche se esse non sono certamente la struttura piú giovane di tutto il panorama marziano.

La scoperta delle incisioni ha resuscitato l'ipotesi che nella storia di Marte possa esserci stata un'epoca in cui quel pianeta somigliava alla Terra. Secondo questa ipotesi, Marte può aver avuto in passato un'atmosfera molto piú densa, contenente vapor d'acqua in quantità tale da consentire la caduta di precipitazioni. Ammessa la caduta di pioggia, le incisioni troverebbero una facile spiegazione. Assai meno facile è invece spiegare perché i canali si sarebbero conservati soltanto in poche aree e perché la topografia piú antica del pianeta non mostri tracce di erosione da parte di acqua corrente. Sarebbe difficile spiegare come l'atmosfera primordiale marziana, probabilmente secca e riducente, possa essersi evoluta, dando luogo a un'altra atmosfera densa e umida che si sarebbe poi nuovamente trasformata nell'atmosfera attuale, rarefatta, secca e costituita quasi completamente di biossido di carbonio. Per di piú, l'atmosfera attuale è mantenuta notevolmente stabile dalla grande quantità di biossido di carbonio solido concentrato nelle regioni polari. Se i canali sono stati creati dalla caduta di pioggia, sembra necessario postulare almeno due miracoli successivi: il primo, la creazione di una atmosfera simile a quella terrestre, che sarebbe perdurata per un periodo abbastanza breve, e l'altro, la sua distruzione.

Un'ipotesi alternativa presenta almeno altrettante difficoltà. Qualcuno ha proposto che l'acqua liquida si sia accumulata in serbatoi sotterranei seguendo poi il destino dell'immobilizzazione e della successiva fusione del permafrost. Questi serbatoi sarebbero stati poi improvvisamente aperti e l'acqua che ne sarebbe uscita avrebbe scavato le incisioni. Queste ultime sono però cosí larghe e profonde che è necessario coinvolgere un volume notevole d'acqua nella loro genesi. Perciò sembra ancora piú difficile considerare le incisioni come frutto di un processo unico, improvviso e a ciclo aperto piuttosto che di un processo a ciclo chiuso come è quello che prevede la presenza di pioggia.

L'origine dei canyon e delle altre incisioni è uno dei misteri principali che sono affiorati con la missione *Mariner 9*. Tenuto conto infatti dell'importanza che l'acqua liquida ha per la vita, la possibilità di ascrivere un certo ruolo all'acqua nella genesi di canyon e incisioni, ha attratto l'interesse da parte di coloro che si occupano del problema.

Infine, ci sono alcune aree su Marte che possono essere descritte dal termine « bacini ». Il piú evidente è la grande struttura circolare, del diametro di piú di 1600 chilometri, chiamata Hellas. Questa struttura è stata osservata da Terra per piú di due secoli. Talora, la sua brillantezza è paragonabile a quella delle calotte polari. *Mariner 7* ha dimostrato che Hellas è in realtà una depressione pressoché priva di strutture. Dall'analisi delle fotografie ad alta risoluzione si è giunti alla conclusione che la superficie di Hellas è stata probabilmente resa uniforme dalla sedimentazione di grandi quantità di polvere là trasportate dal vento. *Mariner 9*, tuttavia, rivelò, proprio mentre la grande tempesta di polvere era sul finire, che Hellas presenta alcune strutture topografiche minori. Questo fatto sembra suggerire che le variazioni nella brillantezza di Hellas possono essere dovute a tempeste di polvere di estensione piú modesta; i primi a giungere a tali conclusioni, sulla base di considerazioni meteorologiche, sono stati C. Sagan della Cornell University e i suoi collaboratori. Cosí Hellas probabilmente agisce a lungo termine come un bacino di raccolta della polvere ma, in determinati casi, quando il vento soffia con particolare violenza, costituisce una sorgente di polvere. L'osservazione da parte di *Mariner 9* di tempeste di polvere di piccole dimensioni e la constatazione che esse possono alterare la luminosità di determinate aree, forniscono ulteriori indicazioni sulla natura di alcune delle strutture variabili che sono state osservate per lunghi anni da terra.

Uno dei risultati piú interessanti della missione *Mariner 7* è stato lo studio con immagini ad alta risoluzione, della vastissima calotta polare presente nell'inverno dell'emisfero meridionale di Marte nel 1969. Le misure di riflettività e di temperatura fornite da uno spettrometro e da un radiometro all'infrarosso collocati sulla sonda spa-

Sulla pagina a fronte è riportata un'immagine che, attraverso una particolare tecnica, sviluppata da J. A. Cutts del Jet Propulsion Laboratory e da J. J. Rennilson dell'Istituto di Tecnologia della California, rappresenta visivamente mutamenti verificatisi sulla superficie di Marte. Essa è stata realizzata con due immagini riprese, a distanza di 37 giorni l'una dall'altra, dalla telecamera grandangolare di *Mariner 9* sulla stessa regione della Depressio Hellespontica. Originariamente entrambe le immagini sono state riprese con lo stesso filtro ma con angolazioni e distanze differenti e sono state poi rettificate attraverso un elaboratore all'Image Processing Laboratory del Jet Propulsion. Per mettere in evidenza i mutamenti intervenuti sulla superficie del pianeta, la prima delle due immagini è stata stampata in verde e la seconda in rosso. Quando esse sono sovrapposte, le zone superficiali che sono diventate piú scure nell'intervallo tra la ripresa della prima e quella della seconda, assumono un aspetto verdastro, mentre quelle che sono divenute piú chiare appaiono piú rossastre. Tutta la zona coperta dall'immagine corrisponde all'incirca alla piú piccola area risolvibile dai telescopi terrestri. Gli astronomi hanno osservato mutamenti stagionali nelle strutture della Depressio Hellespontica: sono la somma di quelli piú particolareggiati visibili in queste immagini.

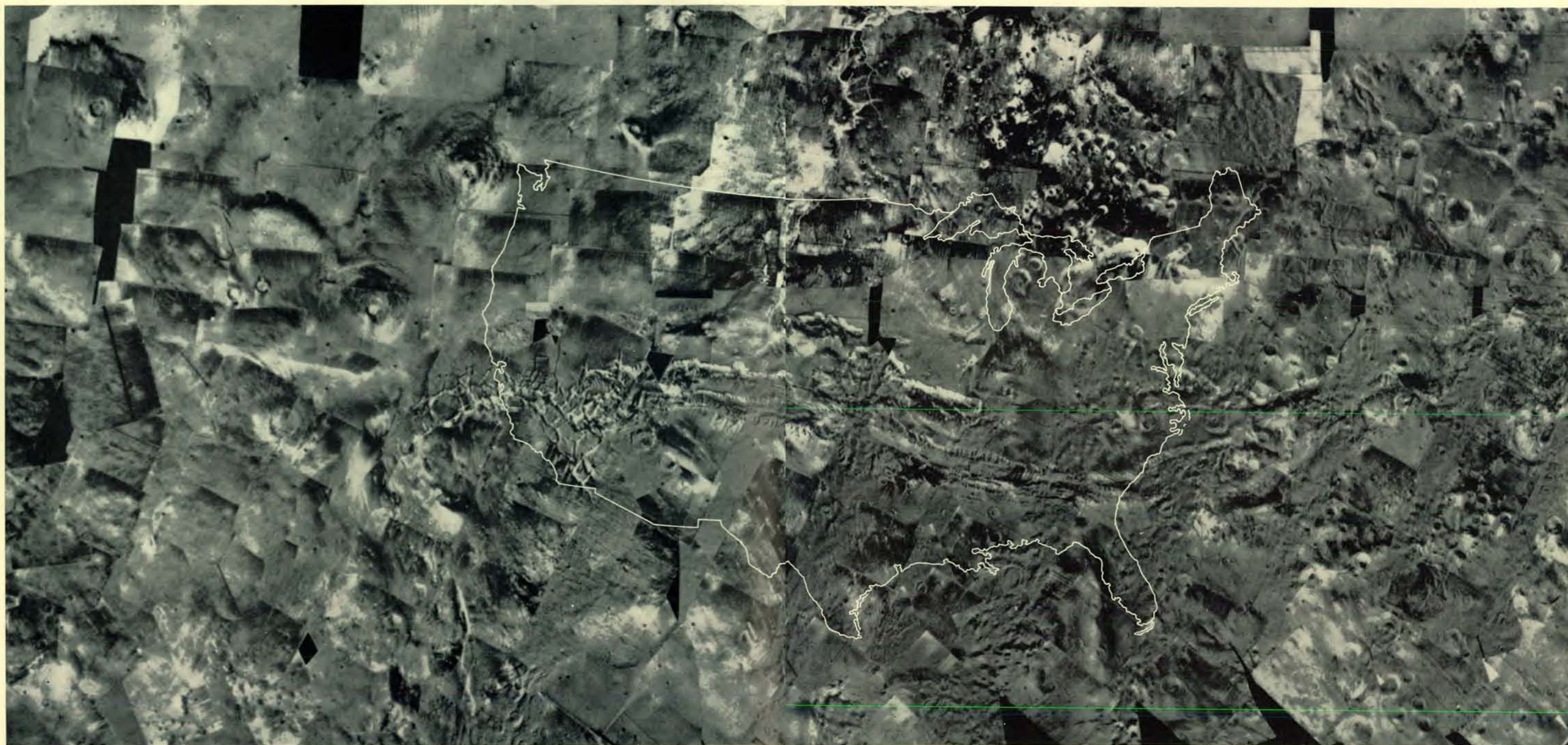
ziale provarono in modo inequivocabile che la calotta polare meridionale è composta di biossido di carbonio solido purissimo, come era stato previsto alcuni anni prima. Le fotografie mostrarono che la copertura glaciale era sottile (in media meno di pochi metri) e che in prossimità del polo sud apparivano anche numerose strutture superficiali insolite.

Quando *Mariner 9* raggiunse Marte era, nell'emisfero meridionale, primavera inoltrata, un periodo ideale per tenere sotto osservazione il ritiro della calotta polare e per esaminare nei particolari le strutture superficiali insolite cui si è accennato. Il ritiro della calotta polare meridionale iniziò come previsto ma, da quel momento, essa mostrò un comportamento anomalo. La calotta polare in ritiro, infatti, curiosamente conservò, nelle sue linee generali, la forma che aveva, anche durante l'estate avanzata, quando il processo di sublimazione di ossido di carbonio avrebbe dovuto raggiungere la punta massima (*si vedano le figure in alto a pagina 73*). Ciò sembra indicare, a mio parere, che, dopo che la vasta calotta annuale di biossido di carbonio solido è sublimata, rimane scoperta una calotta polare residua costituita di normale ghiaccio d'acqua: quest'ultimo, infatti, evapora assai piú lentamente che non il biossido di carbonio; tracce di vapor d'acqua sono presenti nell'atmosfera di Marte.

Le immagini di *Mariner 9* hanno anche mostrato un terreno particolarissimo nella zona polare meridionale; a esso è stato dato il nome di terreno laminato. Anche se la sua forma non è simmetrica, esso si estende su tutta la regione polare meridionale fino a circa 70° di latitudine sud. Il terreno laminato è costituito da strati sottilissimi, alternativamente chiari e scuri, debolmente pendenti e la loro superficie è interessata da un certo rilievo (*si veda l'illustrazione in basso a pagina 73*).

Le sottili lamine sembrano essere raggruppate in pacchetti di 20 o 30 unità a costituire lastre dello spessore di mezzo chilometro o piú e della larghezza che arriva fino a 200 chilometri. La superficie superiore delle lastre ha una struttura a bande. I depositi a lamine sono stati trovati soltanto nelle regioni polari nelle quali il biossido di carbonio forma ogni anno uno strato congelato. Ciò sembra indicare che le laminazioni sono in qualche modo associate con l'andare e venire di sostanze volatili e che esse possono anche trattenere una certa quantità di biossido di carbonio solido o di ghiaccio d'acqua. Dato che le laminazioni sono interessate da pochissimi crateri di impatto, se ne può dedurre che esse sono comparse abbastanza di recente nella storia di Marte.

Il polo nord

La regione polare settentrionale di Marte è diventata osservabile da parte di *Mariner 9* solo a missione abbastanza inoltrata, in conseguenza del gra-

Un fotomosaico ottenuto dalla riunione di diverse decine di immagini grandangolari di *Mariner 9*; esse sono state elaborate e riunite da R. Batson dell'US Geological Survey di Flagstaff in Arizona. La zona coperta da questa panoramica costituisce solo una parte della fascia equatoriale ed è indicata nella carta geologica di Marte alle pagine 64 e 65. Per dare un'idea delle dimensioni dei grandi canyon che si estendono parallelamente all'equatore per 4000 chilometri è stato qui sovrapposto il profilo degli USA. I canyon, strutture assai piú grandi di qualunque analogo terrestre, hanno una larghezza media di 100 chilometri e una profondità di oltre 6 chilometri. All'estremità sinistra del fotomosaico è ben visibile il gigantesco edificio della Nix Olympica e, immediatamente alla sua destra, la serie di vulcani un po' piú piccoli che formano la dorsale di Tharsis. Quando sono state riprese le immagini, il Sole illuminava il paesaggio da sinistra in basso: le ombre si estendono verso il lato superiore destro del fotomosaico.

duale mutamento dell'illuminazione associato con il cambio di stagione e con il sollevarsi della nebbia caratteristica che si sviluppa su ciascun polo durante la stagione invernale. Le strutture quasi circolari caratteristiche del terreno laminato sono state trovate attorno al polo nord in quantità ancora maggiore che non attorno al polo sud. Si possono infatti osservare 20 o 30 lastre singole, fra loro parzialmente sovrapposte in un modo che ricorda la sistemazione assunta da una manciata di gettoni da poker gettata sul tavolo.

Il fatto che il terreno laminato e le lastre circolari siano presenti sia nella regione polare settentrionale sia in quella meridionale indica, al di là di ogni ragionevole dubbio, che la loro formazione deve in qualche modo essere associata con la periodica deposizione e sublimazione di sostanze volatili.

Michael C. Malin, dell'Istituto di Tecnologia della California, e io abbiamo fatto l'ipotesi che la distribuzione delle lastre circolari e la loro sovrapposizione possa essere spiegata con mutamenti nell'inclinazione dell'asse di Marte. Abbiamo supposto che l'asse di rotazione del pianeta si sia spostato durante le ultime decine di milioni d'anni come conseguenza dell'attività di correnti di convezione profonde entro il mantello del pianeta e probabilmente associate con lo sviluppo di vulcani nelle zone equatoriali. Mentre l'asse di rotazione si spostava, le lastre laminate si sono deposte concentricamente attorno a ciascuna delle posizioni assunte dai poli.

Quest'ipotesi è perlomeno in accordo con le informazioni sulla distribuzione della gravità di Marte dedotte dalle variazioni dell'orbita di *Mariner 9*. Il pianeta mostra anomalie gravitazionali che sembrano suggerire differenze di densità in profondità del tipo che si riscontrerebbe nel caso fossero attive correnti di convezione profonde. Per di piú, vi è una notevole correlazione tra le anomalie gravitazionali e la posizione dei vulcani equatoriali.

L'alternarsi regolare delle lamine e delle lastre stesse sembra indicare che esse sono anche connesse in qualche modo con variazioni periodiche del clima di Marte. In collaborazione con William Ward e Sze Yeung, ho studiato le variazioni teoriche dell'orbita di Marte su un certo periodo di tempo. Le perturbazioni nell'orbita causate da altri pianeti, analizzate anni fa da D. Brouwer e G. M. Clemence, alterano l'eccentricità dell'orbita in un modo che appare essere abbastanza in accordo con la nostra ipotesi. L'eccentricità dell'orbita di Marte varia da 0,004 (cioè da quasi circolare) a 0,141, mentre il suo valore attuale è 0,09 (*si veda la figura in alto a pagina 76*).

La conseguenza di questa variazione nell'eccentricità è una variazione nella quantità annuale media di luce solare che raggiunge i poli del pianeta insieme con una variazione molto maggiore nel flusso solare massimo nei momenti in cui il pianeta è piú prossimo al Sole (*si veda la figura in basso a pagina 76*). Anche se la variazione nella quantità media di energia raggiante che raggiunge i poli riguarda soltanto una percentuale minima, in determinate circostanze, essa è sufficiente a causare una variazione ciclica nella crescita e nella sublimazione delle calotte polari permanenti di biossido di carbonio solido. Supponendo che le tempeste di polvere depositino regolarmente un manto di polvere durante le fasi di sublimazione, potrebbero formarsi sottili lamine del tipo che in realtà è stato osservato. Le lastre stesse potrebbero a loro volta corrispondere a una periodicità dell'ordine dei due milioni di

La zona del polo sud in differenti stagioni, ripresa da *Mariner 7* nell'agosto 1969 (*a sinistra*) e da *Mariner 9* nel novembre 1971 (*a destra*). L'area ritratta dalle due immagini di *Mariner 9* è indicata dal riquadro sovrapposto al mosaico di *Mariner 7*. Il polo sud geografico è indicato in entrambi i mosaici da un punto bianco. Nell'agosto 1969 l'emisfero meridionale di Marte si trovava in inverno e la calotta di biossido di carbonio solido (ghiaccio secco) era prossima al suo massimo sviluppo. La porzione destra del mosaico è scura perché si trova nell'ombra. Quando 27 mesi piú tardi furono riprese le immagini di *Mariner 9*, la calotta polare stava invece dissolvendosi rapidamente, con l'avvento della stagione primaverile nell'emisfero meridionale di Marte.

Il graduale dissolversi della calotta polare meridionale in tre immagini riprese da *Mariner 9*. Le immagini originali sono state elaborate, rettificate stereograficamente e stampate alla stessa scala. Ciascuna di esse mostra l'aspetto che la calotta polare aveva dopo che la sonda si trovava in orbita da 14, 36 e 94 giorni. La prima immagine a sinistra è ancora un poco nebbiosa poiché era ancora in atto la grande tempesta di polvere. Si può vedere, tuttavia, che la forma generale della calotta è cambiata in modo significativo nelle tre settimane intercorse tra la prima e la seconda ripresa ed è invece cambiata assai poco nelle successive otto settimane (tra la seconda e la terza), quando la sublimazione della calotta avrebbe dovuto essere al suo massimo.

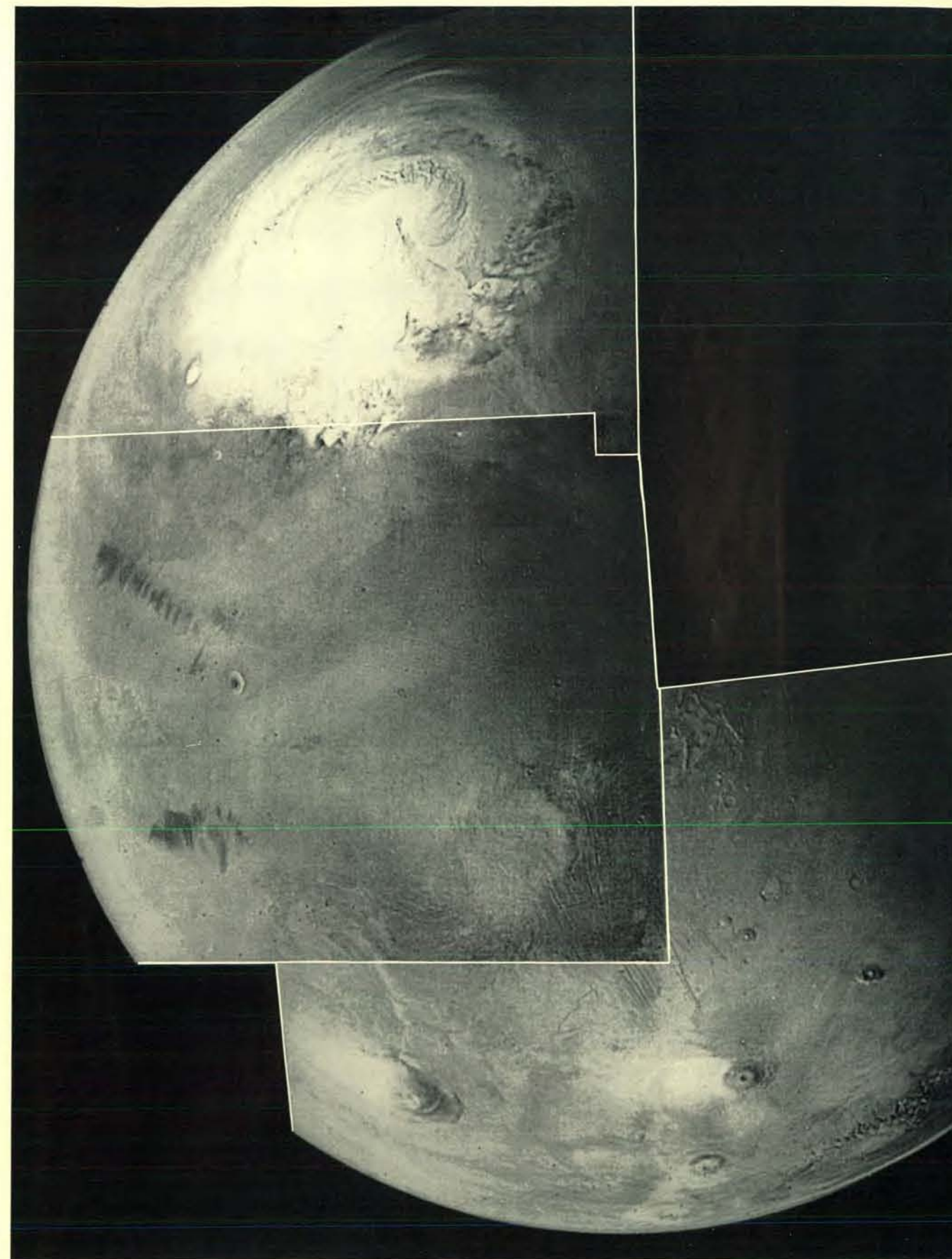
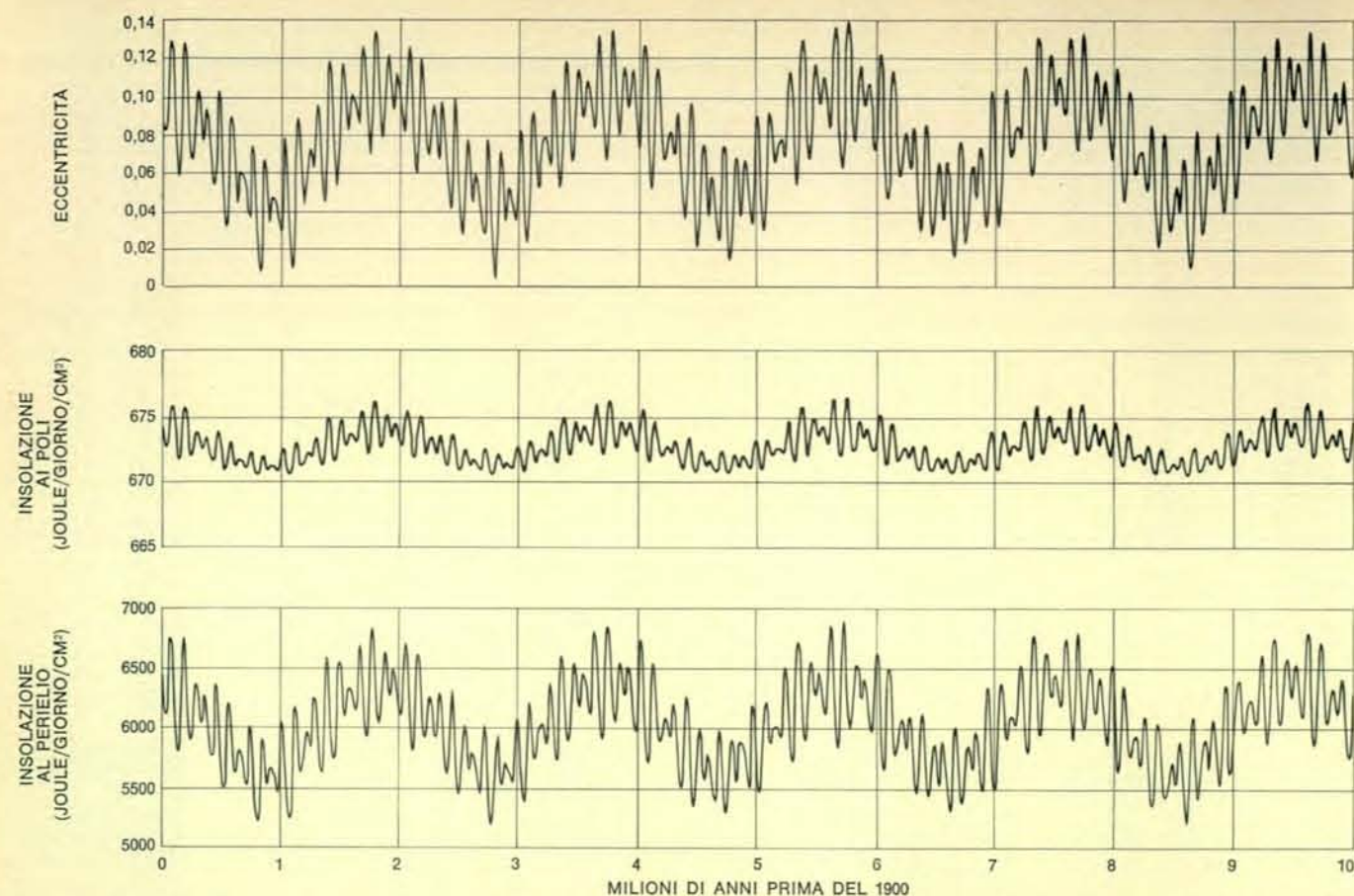
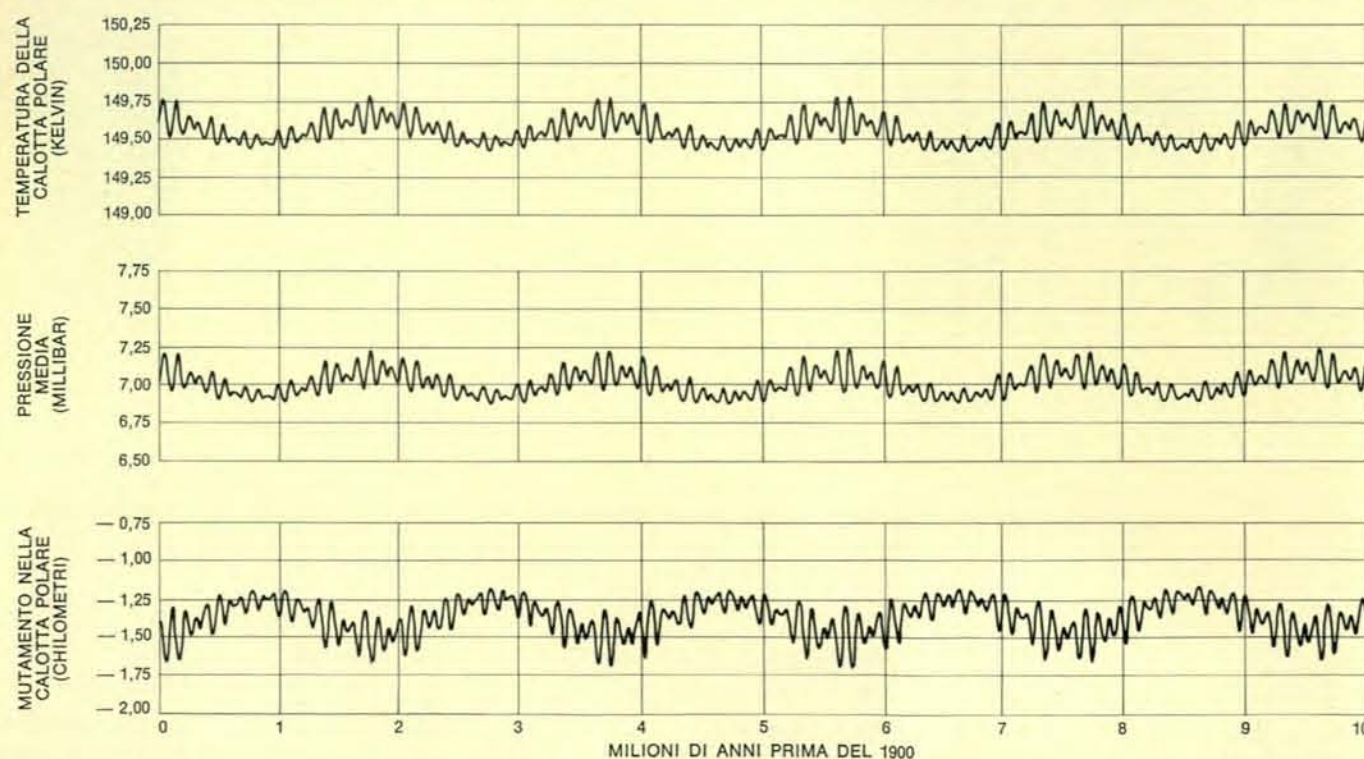
Un mosaico della calotta polare meridionale residua ottenuto da L. A. Soderblom, dell'US Geological Survey al Centro di astrogeologia di Flagstaff in Arizona, con una serie di immagini ravvicinate riprese tra il 58° e il 94° giorno di volo orbitale. La calotta polare residua, che ha un diametro di 300-350 chilometri, è centrata a circa 200 chilometri di distanza dal vero polo sud (*puntino bianco*), a circa 45° di longitudine ovest. Le ombre grigie sono state elaborate in modo tale da far risaltare i particolari nella regione coperta dalla calotta; per questa ragione le superfici libere dal ghiaccio appaiono nere.

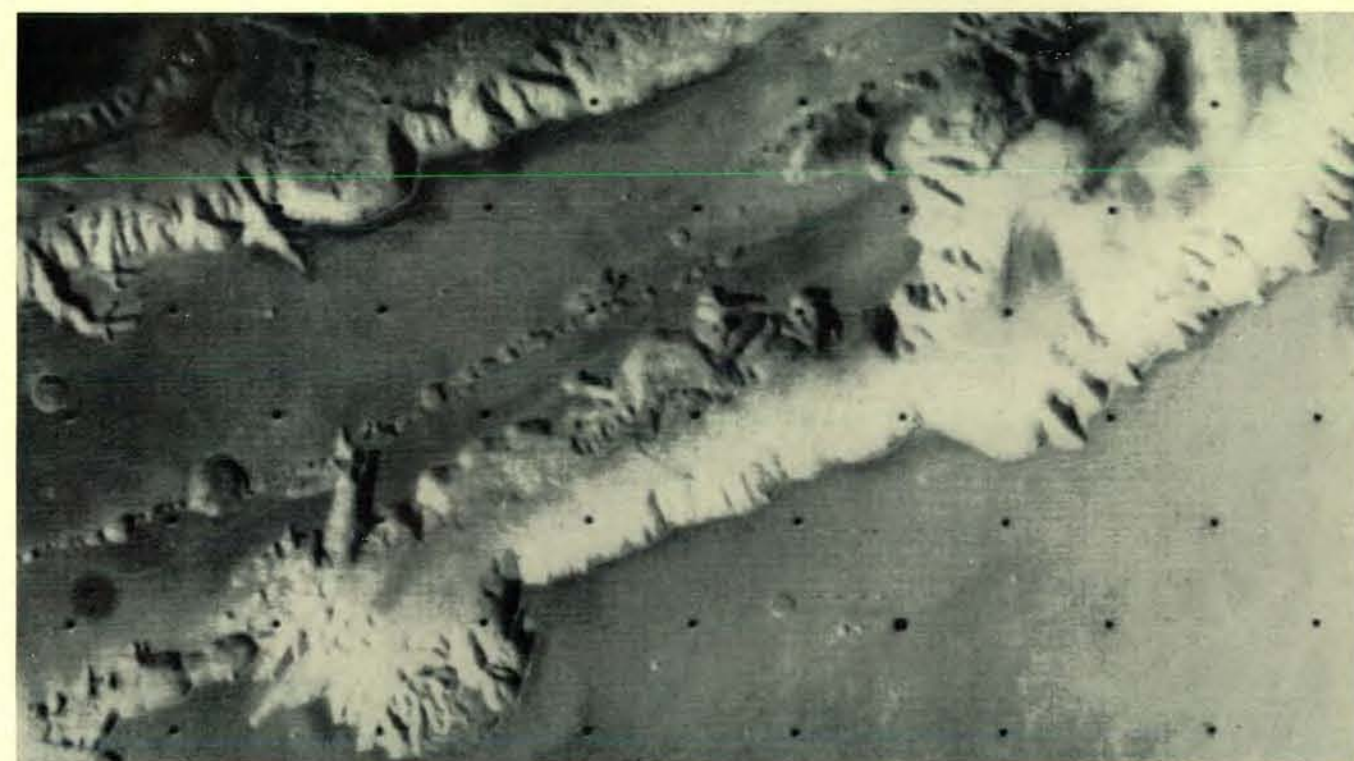
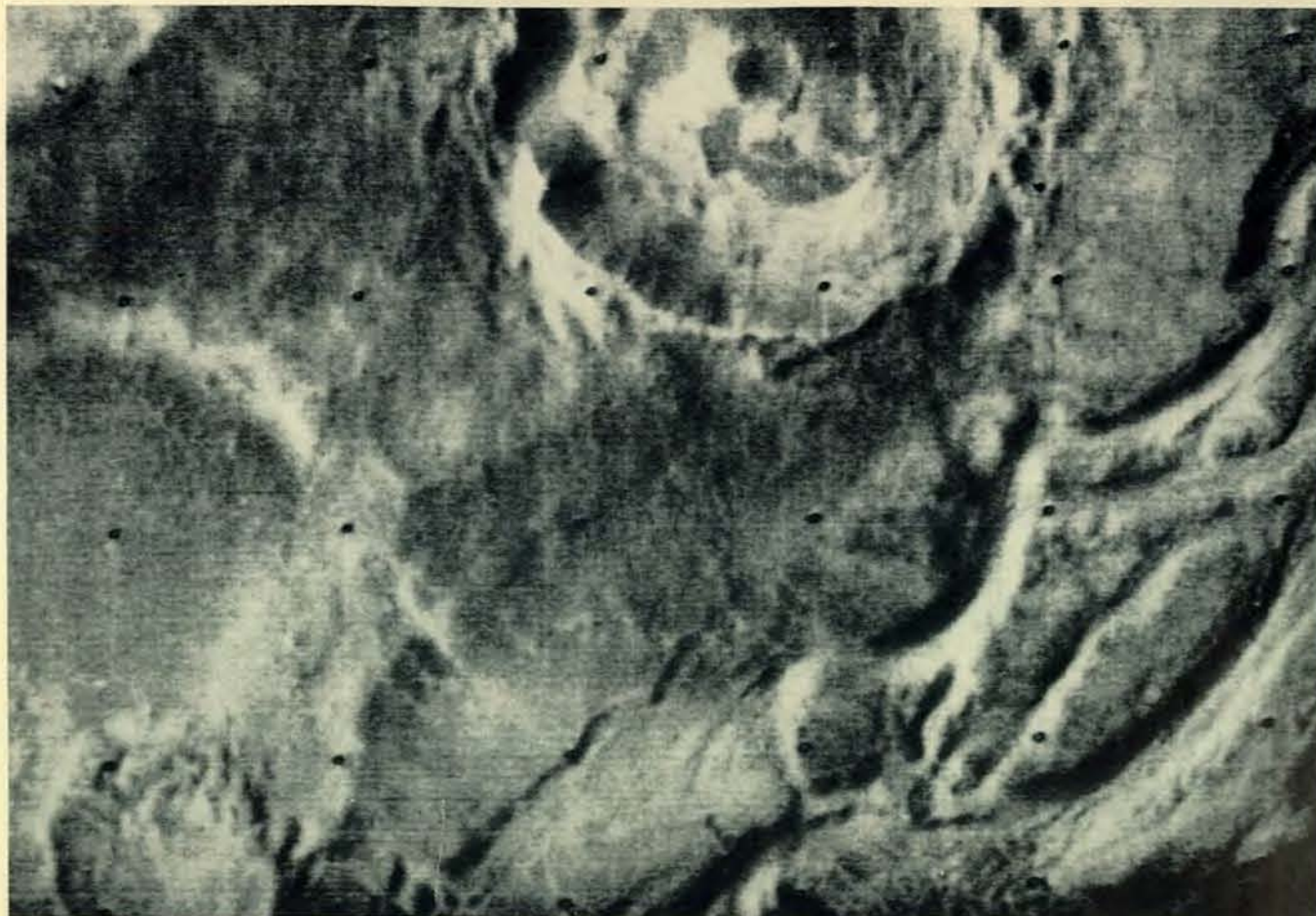
Il terreno laminato della regione polare settentrionale: l'autore ipotizza che la distribuzione delle lastre circolari e la loro parziale sovrapposizione siano prove che l'inclinazione dell'asse di rotazione di Marte è variata negli ultimi 100 milioni di anni.

anni. Perciò il terreno laminato sembra riflettere strettamente le fluttuazioni nel flusso medio di energia raggiante che raggiunge il pianeta: fluttuazioni a breve scadenza, con una periodicità di circa 90 000 anni, e fluttuazioni a lunga scadenza, con una periodicità di circa 2 milioni di anni. Di conseguenza, dato che nell'emisfero settentrionale sono visibili 20 o 30 lastre, il terreno laminato costituisce in questo caso la testimonianza di un periodo corrispondente all'incirca agli ultimi 100 milioni di anni. Un'ipotesi alternativa sulla genesi del terreno laminato è quella che ne fa una struttura soprattutto erosionale piuttosto che non deposizionale.

Evoluzione dell'atmosfera

Se tutte le laminazioni che si trovano nelle zone polari si sono accumulate soltanto nelle ultime poche centinaia di milioni di anni, cioè in non piú del 5 % della storia di Marte, cosa è accaduto prima? E qui si incontra una difficoltà di fondo se si cerca di applicare quello che è considerato uno dei principi base della geologia, il principio di J. Hutton, secondo il quale « il presente è la chiave del passato ». Infatti, sia che si guardi alle zone vulcaniche, a quelle dei canyon, alle incisioni o alle laminazioni polari, tutto sembra essere il risultato di un'attività intensa e di notevoli cambiamenti avvenuti nella parte piú recente del passato geologico di Marte. Queste considerazioni mi hanno indotto a pensare se, per caso, non sia possibile che l'atmosfera di Marte, cosí come la conosciamo, non possa essere stata acquisita abbastanza di recente. Malin e io stiamo presentando questa idea sotto forma di ipotesi di lavoro: può darsi che Marte non abbia avuto del tutto atmosfera oppure ne abbia avuta soltanto una molto esile e priva di importanza lungo tutto il periodo intermedio della sua storia, durato forse alcuni miliardi di anni. Presumibilmente Marte aveva una atmosfera primordiale formatasi nel corso della accrezione del pianeta; se però era costituita soprattutto di idrogeno e di metano, essa deve essere stata perduta abbastanza presto.

Riteniamo che una porzione significativa della attuale atmosfera di Marte sia stata liberata durante la formazione della Nix Olympica e degli altri tre vulcani della zona di Tharsis. L'esistenza di vasti accumuli di materiale roccioso e altri indizi di sedimentazione e di un vulcanismo un poco anteriore e piú esteso, suggeriscono che, in un periodo geologico abbastanza re-

L'emisfero settentrionale di Marte in tre immagini, riprese da 13 700 chilometri a 84 secondi l'una dall'altra, facendo ruotare la telecamera grandangolare del *Mariner 9*. L'immagine in basso mostra chiaramente la Nix Olympica, i vulcani della dorsale di Tharsis e, nell'angolo in basso a destra, il canyon che si estende appena al di sotto dell'equatore. La parte settentrionale del pianeta, a nord del 50° parallelo, appare oscurata da nubi di ghiaccio d'acqua o di cristalli di biossido di carbonio.

Le particolarissime lastre circolari e le laminazioni osservate nelle regioni polari possono, secondo un'ipotesi sviluppata dall'autore e dai suoi collaboratori, essere spiegate con mutamenti nell'orbita di Marte intervenuti nel corso degli ultimi 10 milioni di anni. L'eccentricità dell'orbita di Marte (*curva in alto*) è variata da 0,004 (orbita quasi circolare) a 0,141. Questo avrebbe comportato il mutamento nella quantità di energia solare che avrebbe in tal caso raggiunto i poli (*curva al centro*) e contemporaneamente nell'energia massima che avrebbe raggiunto il pianeta quando esso si trovava più vicino al Sole (*curva in basso*).

I mutamenti nell'energia solare su Marte, dovuti a variazioni nell'eccentricità dell'orbita del pianeta, potrebbero produrre variazioni cicliche nel deposito e nella sublimazione del biossido di carbonio delle calotte polari. La curva in alto mostra i mutamenti a lunga scadenza nella temperatura del biossido di carbonio solido (ghiaccio secco) ai poli. La curva al centro mostra variazioni nella pressione atmosferica media alla superficie del pianeta, così come potrebbe risultare dalle variazioni nell'insolazione. La curva in basso mostra i numerosi cambiamenti che ne conseguirebbero nello spessore della calotta polare permanente.

cente, notevoli quantità di sostanze volatili siano state emesse dall'interno del pianeta. Quindi potrebbe essere accaduto che Marte, una volta maturato abbastanza come pianeta da raggiungere ovunque le condizioni per la liberazione dei volatili, abbia cominciato a sviluppare una sua atmosfera duratura. Questa, a sua volta, ha determinato la formazione dei terreni laminati e ha fornito un meccanismo di trasporto eolico e di erosione, sufficiente a formare incisioni e grandi canyon. Secondo questa ipotesi Marte è ancora ben lontano dal raggiungere lo stato stazionario

In queste due immagini della regione di Coprates, presso l'equatore marziano, è documentata la fine della tempesta di polvere. L'immagine in alto, ripresa quando *Mariner 9* era in orbita da 41 giorni, non reca quasi traccia del tormentato canyon che compare invece nell'immagine in basso, ripresa quando *Mariner 9* era in orbita da 80 giorni. Coprates è da lungo tempo noto agli astronomi come una struttura che cambia in modo imprevedibile di luminosità. Ora è chiaro che il canyon appare più brillante delle regioni circostanti verso la fine di una tempesta di polvere, quando cioè l'atmosfera nel canyon è ancora colma di particelle in sospensione altamente riflettenti, mentre l'atmosfera dell'altipiano circostante risulta ormai quasi limpida.

Intreccio di incisioni adiacente a un cratere di impatto del diametro di 20 chilometri; è questo un esempio del tipo di strutture che suggerisce l'erosione da parte di un qualche fluido sulla superficie di Marte. Se in realtà l'agente è stato l'acqua, è difficile capire perché le tracce di erosione sono confinate soltanto in ristrette regioni della superficie totale del pianeta.

in cui l'erosione dovrebbe essere compensata dalla modificazione della superficie, necessaria allo sviluppo di uno spettro abbastanza ampio di strutture morfologiche. Nel quadro della nostra ipotesi di lavoro si può anche immaginare che durante i primi stadi nello sviluppo dell'attuale atmosfera marziana, prima che le calotte polari si instaurassero definitivamente intrappolando parte del biossido di carbonio, fosse caduta sulla superficie del pianeta dell'acqua in quantità sufficiente a scorrere entro le incisioni in condizioni del tutto particolari e non ricorrenti. Questa soluzione almeno evita il problema di supporre che siano avvenuti due miracoli l'uno dopo l'altro: essa ne richiede uno solo se riteniamo che l'acqua liquida sia realmente necessaria a spiegare la genesi delle incisioni.

L'ipotesi di un'atmosfera giovane può anche aiutare a spiegare come mai le strutture scure permanenti (per esempio, Sinus Meridiani) possano conservarsi malgrado le frequenti tempeste di sabbia di estensione planetaria. Ancora una volta Marte non ci sembra essere in uno stato stazionario, anche se altri ricercatori non condividono il nostro punto di vista. Secondo la nostra ipotesi, le strutture scure possono essere l'affioramento di materiali superficiali piú antichi non ancora interessati dalla degradazione chimica connessa con la nuova atmosfera. Di fatto, sembra esservi una qualche correlazione tra le strutture scure permanenti e le zone in cui si trovano i crateri piú antichi.

Altri ricercatori che hanno analizzato le immagini di *Mariner 9*, come Sagan e W. K. Hartmann, hanno sviluppato un'ipotesi abbastanza diversa sulla storia di Marte. La natura delle zone antiche a crateri li ha indotti a pensare che un lungo periodo di erosione atmosferica abbia preceduto gli eventi spettacolari del piú recente passato. Cosí il concetto di un pianeta Marte simile alla Terra non si è per nulla estinto. Ciononostante, le idee sulla storia geologica di Marte stanno mutando rapidamente alla luce dei dati accumulati dal pieno successo della missione *Mariner 9*. Qualche interpretazione intermedia potrà fornire un quadro piú organico di tutte le osservazioni.

Vita su Marte?

I risultati attuali della missione *Mariner 9*, tuttavia, mi inducono a un modo di vedere assai differente da quello dei primi astronomi che ritenevano che Marte fosse una volta simile alla Terra e che sia ora divenuto null'altro che un corpo asciutto e morto. A mio parere Marte sta probabilmente cominciando proprio ora a divenire simile alla Terra, sviluppando un'atmosfera durevole. « Proprio ora » è una locuzione difficile da definire in modo quantitativo, poiché le età assegnate ai crateri sulla base del flusso di meteoriti sono ancora assai incerte. Secondo me, la formazione dell'attuale atmosfera, ammesso che essa si sia formata come è stato detto piú sopra, è avvenuta nell'ultimo quarto della storia di Marte e, comunque, di certo non prima dell'ultima metà. Se questa ipotesi di lavoro dovesse essere largamente accettata, ne conseguirebbe necessariamente un certo pessimismo sulla possibilità che le condizioni sul pianeta siano state in passato favorevoli alla comparsa di forme semplici di vita. Se Marte, invece, è simile alla Luna, ed è quindi rimasto privo di una atmosfera significativa per gran parte della sua storia, e se la massima quantità di acqua disponibile sulla sua superficie è stata appena sufficiente a scavare qualche incisione, è altamente improbabile che vi sia mai stato un accumulo superficiale di acqua liquida sufficiente a permettere lo sviluppo accidentale di qualche forma di vita da materiale organico e biologico. D'altro canto, i sostenitori della vita su Marte la pensano diversamente e sottolineano il fatto che, se dell'acqua è stata in qualche momento disponibile sulla superficie marziana, essa ha potuto fornire un ambiente favorevole allo sviluppo della vita. Ovviamente questa disputa non può essere risolta dal tipo di informazioni raccolte da *Mariner 9*: la risposta deve attendere le analisi chimiche e mineralogiche accurate del suolo del pianeta.

Una valle sinuosa della lunghezza di 400 chilometri e della larghezza di oltre 6 chilometri, situata a 29° di latitudine sud e 40° di longitudine ovest, nella regione del Mare Erythraeum. La struttura assomiglia all'andamento di un fiume a meandri oppure a uno dei rills lunari che possono essere stati scavati da una colata lavica. L'origine di queste strutture non è chiara.

Contemporaneamente alla missione di *Mariner 9*, era in atto anche una ambiziosa missione sovietica il cui scopo era di deporre sulla superficie di Marte una sonda che eseguisse alcune analisi della superficie. Sfortunatamente, tale sonda, *Mars 3*, cessò di funzionare poco dopo aver raggiunto la superficie e, a quanto sembra, non ha trasmesso alcuna informazione utile. Occorre attendere che l'URSS ripeta la missione negli ultimi mesi del 1973 e che si possano vedere le immagini ravvicinate della superficie e i risultati di alcune semplici analisi chimiche.

Si spera che per il 1976 saranno disponibili le informazioni ottenute da un complesso veicolo di atterraggio su Marte, sviluppato nel quadro del programma *Viking* e forse da sonde russe piú avanzate. La capsula *Viking* è stata progettata non solo per cercare direttamente composti organici, ma anche per compiere alcune determinazioni semplici, ma importanti, della composizione fondamentale dei minerali superficiali. Tali misure forniranno uno strumento importante per ricostruire l'evoluzione chimica passata dei minerali superficiali e quindi per sapere anche se essi hanno o meno reagito chimicamente con l'acqua.

La missione *Viking* finirà col dirci finalmente se su Marte esiste o è esistita la vita? Ho i miei dubbi. Ritengo infatti che la difficoltà di ottenere una risposta inequivocabile a questo problema sia talmente grande da essere al di là delle possibilità anche di una missione ambiziosa come quella dei *Viking*. Ritengo che la risposta dovrà attendere il ritorno sulla Terra di un campione di suolo marziano, raccolto probabilmente da una sonda automatica, perché possa essere sottoposto alle raffinate indagini possibili solo nei laboratori terrestri. A mio avviso, l'URSS, avendo già dimostrato con la Luna di avere sviluppato la tecnologia necessaria a raccogliere e riportare a Terra campioni di suolo con sonde automatiche, potrebbe essere in condizioni di ripetere l'operazione con Marte attorno al 1980; ciò naturalmente a condizione che l'URSS continui a riservare all'esplorazione automatica del cosmo la stessa priorità che in passato. Gli USA non hanno piani ambiziosi di esplorazione di Marte se si eccettua la missione *Viking*. *Mariner 9* verrà, così, ricordato a lungo come una delle punte avanzate dell'esplorazione americana di Marte.

Cristalli di neve

Il classico cristallo di neve esagonale finemente ramificato è solo una delle pressoché infinite varietà di forme. Ciascuna forma dipende dalla modalità di crescita del cristallo di neve nella nube

di Charles e Nancy Knight

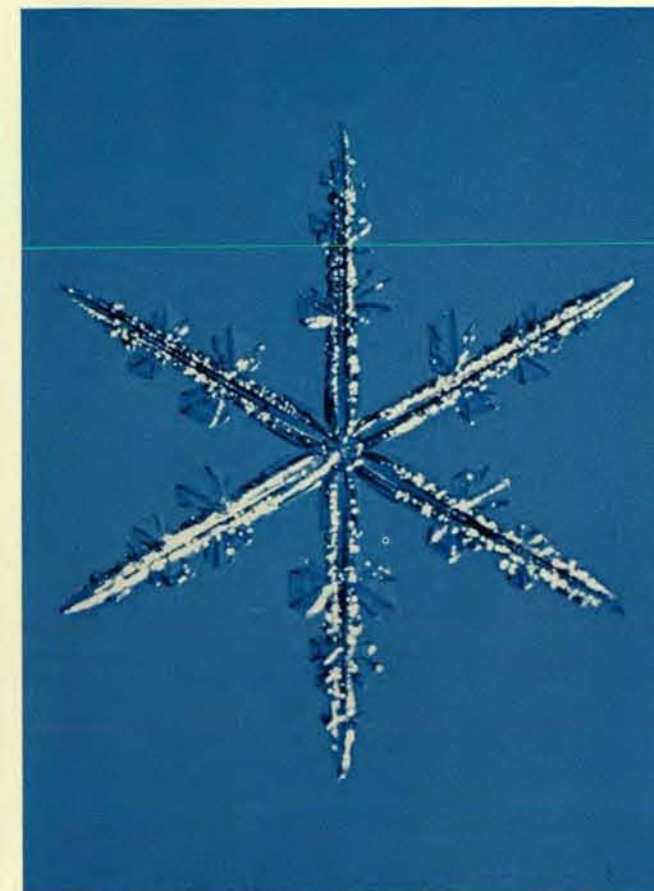
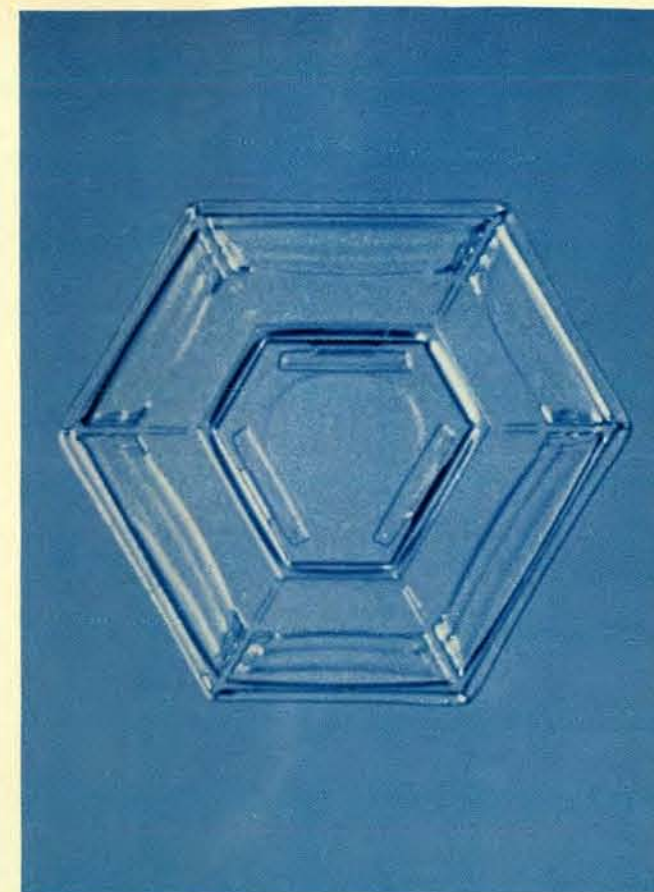
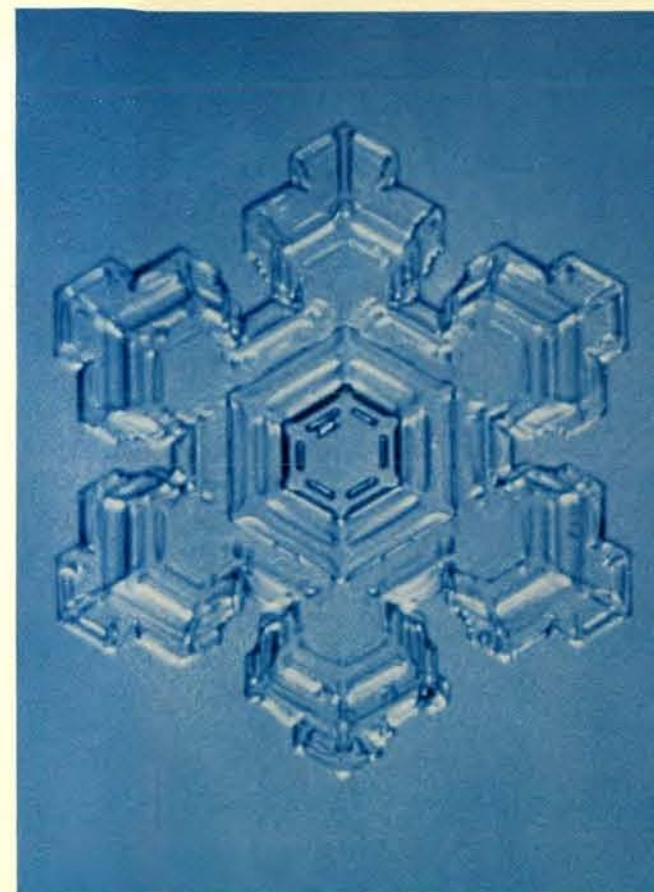
La comune forma esagonale di un cristallo di neve riflette la disposizione delle molecole d'acqua nella struttura del ghiaccio. Un esame approfondito dei cristalli naturali di neve rivela però che se si eccettua la forma esagonale di base, i cristalli sono ben lungi dall'essere uniformi. Se si raccolgono cristalli di neve fresca e si esaminano al microscopio, vi sono buone probabilità che molte forme non siano per niente comuni. Alcune sono lamelle piatte; altre lunghi aghi; altre sono così complesse e apparentemente irregolari che quasi non è possibile descriverle. La costituzione, o il particolare aspetto del cristallo, dipende dalle condizioni in cui si è formato. Prelevando cristalli di neve nel loro ambiente naturale ed esaminandoli è possibile individuare queste condizioni.

Poiché la neve si forma nelle nubi, consideriamo prima di tutto alcune caratteristiche generali delle nubi. Le nubi si formano quando l'aria si raffredda sotto la temperatura alla quale è satura di vapore d'acqua. Il raffreddamento è generalmente causato dalla espansione cui è soggetta l'aria quando da una regione a pressione più alta sale verso una regione a pressione più bassa. Vi sono tre tipi comuni di risalita dell'aria. Il primo è la convezione e si ha quando aria relativamente calda, leggera e umida sale in pennacchi in una zona di aria più fredda, più densa e più secca. La velocità di risalita può essere di molti metri al secondo. Il secondo tipo di risalita, è provocato da movimenti atmosferici su larga scala. In questo caso l'ascensione è lenta con una velocità tipica di pochi centimetri al secondo; l'aria può salire però su vaste aree e per la durata di tutto un giorno. Il terzo tipo di risalita si ha quando venti orizzontali incontrano dei rilievi montuosi. Questi tre fenomeni possono agire singolarmente oppure combinati tra di loro.

Quando la temperatura alla quale l'aria raggiunge la saturazione è superiore al punto di congelamento, ossia 0 °C, si forma una nube di goccioline d'acqua. La base della nube è piatta e l'altezza a cui si trova è il livello di condensazione dell'acqua. Quando la temperatura alla quale l'aria diventa satura è inferiore a 0 °C, caso tipico delle nubi che producono neve, vi sono due livelli di condensazione: uno per il ghiaccio e uno per l'acqua sovraraffreddata, ossia l'acqua che rimane liquida al di sotto di 0 °C. Questa situazione si verifica perché l'acqua sovraraffreddata ha sempre una tensione di vapore superiore a quella del ghiaccio alla stessa temperatura. (La tensione di vapore è la pressione di equilibrio tra la fase gassosa e la fase liquida o solida di una sostanza.) Di conseguenza quando i due livelli di condensazione coesistono, il livello di condensazione del ghiaccio è sempre il più basso, ossia è il primo livello che l'aria raggiunge quando sale.

Poiché l'aria che sale e si raffredda raggiunge prima la saturazione rispetto al ghiaccio, ci si dovrebbe aspettare di trovare nubi di ghiaccio a fondo piatto con basi a quel livello. Questo però normalmente non accade. La aria fredda può conservarsi soprasatura rispetto al ghiaccio senza che si formino cristalli, per la stessa ragione per cui l'acqua sovraraffreddata non necessariamente gela. In entrambi i casi sono necessarie particelle estranee che fungono da nuclei attorno ai quali il ghiaccio si cristallizzi e tali particelle sono in natura relativamente scarse. Normalmente quindi una nube non si forma al primo livello di condensazione. Anche il vapore d'acqua ha bisogno di nuclei solidi per condensarsi in gocce ma il tipo di particelle necessarie è molto più comune e generalmente assai abbondante in natura. L'aria ascendente può oltrepassare il livello di condensazione del ghiaccio senza che si formi una nube di ghiaccio; una volta però che l'aria raggiunge il livello di condensazione dell'acqua, si forma una nube composta di piccole gocce di acqua sovraraffreddata.

Nonostante che la natura sia generalmente povera di nuclei di congelamento sembra che quelli che ci sono siano più efficaci nel favorire il congelamento delle gocce d'acqua sovraraffreddata che non nel formare direttamente ghiaccio dal vapore. In parte le gocce sovraraffreddate che si trovano nella nube contengono o vengono a collisione con nuclei di congelamento e si trasformano in ghiaccio. A questo punto diventa rilevante la differenza tra la tensione di vapore del ghiaccio e quella dell'acqua sovraraffreddata. Quando gocce di acqua sovraraffreddata e particelle di ghiaccio coesistono, le molecole di vapore che si trovano lungo la superficie delle goccioline d'acqua si diffondono verso il ghiaccio. La goccia d'acqua comincia a evaporare e il ghiaccio si accresce per condensazione diretta dal vapore.

Mentre cadono, i cristalli crescono, ma poiché cadono in correnti ascendenti può darsi che in realtà si innalzino rispetto al suolo, soprattutto nelle nubi convettive. Le nubi variano nel tempo e la storia della crescita di un singolo cristallo di neve può avere in pratica qualsiasi corso. Probabilmente l'abito più comune di un singolo cristallo è quello tabulare dendritico: ne risulta un cristallo piatto con rami delicati spesso considerato come il fiocco di neve tipico. I più belli di questi cristalli di neve provengono da correnti ascendenti moderate in cui la maggior parte della crescita del cristallo avviene mentre esso cade al di sotto del livello

I cristalli tabulari di neve hanno la comune forma esagonale. I due cristalli in alto sono tabulari; i due in basso dendritici. Queste forme non sono riproduzioni ma fiocchi originali. Sono stati raccolti da Teisaku Kobayashi dell'Istituto di scienza delle basse temperature di Sapporo in Giappone e sono stati fotografati su un vetrino al microscopio sia a luce trasmessa sia a luce riflessa.

di condensazione dell'acqua. Le nevi leggere delle notti calme e fredde sono spesso di questo tipo. Nelle nevi piú pesanti le singole particelle generalmente si aggregano in grandi e soffici fiocchi di neve. Questi aggregati grandi e leggeri consistono di cristalli dendritici, mentre i fiocchi un po' piú pesanti e a caduta piú veloce possono essere grappoli di cristalli a forma di proiettile.

Da lungo tempo la forma esagonale dei cristalli di neve costituisce oggetto di attenzione. Giovanni Keplero scrisse un saggio sulle forme dei cristalli di neve nel XVI secolo. Attualmente la forma esagonale è spiegata in termini di disposizione delle molecole all'interno dei cristalli. La velocità con cui le molecole d'acqua trovano i punti di legame sulla superficie del cristallo dipende dalla struttura della superficie. A sua volta la struttura della superficie dipende dal suo orientamento rispetto al reticolo esagonale delle molecole di ghiaccio da cui è costituito (*si veda l'illustrazione in basso nella pagina a fronte*). Nonostante che i particolari del processo di crescita del cristallo siano difficili da studiare a causa delle piccole dimensioni della scala, è chiaro che alcuni orientamenti della superficie del cristallo possono catturare le molecole d'acqua nel reticolo cristallino piú velocemente di altri. Il processo dà origine alle forme cristalline la cui simmetria esterna esagonale riflette la simmetria interna dei legami molecolari.

Il metodo piú semplice per studiare i cristalli di neve è quello di farne un duplicato facendoli cadere in un sottile strato di una soluzione diluita di una resina sintetica. Il solvente evapora rapidamente, lasciando un sottile calco di plastica del cristallo di neve. Questa tecnica funziona meravigliosamente bene con i cristalli tabulari, mentre funziona molto male quando i cristalli sono complessi e tridimensionali. Ciononostante è comoda per esaminare i cristalli di neve; infatti essi vengono osservati con un microscopio a basso potere risolvente ed è molto piú semplice lavorare col microscopio a temperatura ambiente che non sotto zero.

Per poter studiare i fiocchi di neve piú complessi bisogna sopportare qualche disagio. La neve può essere raccolta ed esaminata al microscopio in un ambiente non riscaldato o all'esterno. Poiché la neve evapora piuttosto rapidamente anche a temperature inferiori a 0 °C, questo esame deve essere veloce per assicurarsi che vengano fedelmente registrati tutti i particolari. Un modo per conservare i cristalli senza duplicarli è quello di raccoglierli in un liquido freddo che non solubilizzi il ghiaccio. Con questo sistema possono essere studiati e fotografati a piacere; l'ambiente tuttavia deve essere mantenuto freddo. È il metodo che abbiamo scelto per il nostro lavoro al National Center for Atmospheric Research. Il liquido che usiamo è l'esano, ma il cherosene o la benzina senza piombo sarebbero ugualmente soddisfacenti. Il nostro procedimento ha l'ulteriore vantaggio di ridurre il contrasto tra il cristallo e lo sfondo perché l'indice di rifrazione del liquido non è molto diverso da quello del ghiaccio e quindi i particolari delle particelle di neve piú complesse sono molto piú chiari.

Il metodo tradizionale di raccolta dei cristalli di neve consiste nel catturarli su un pezzo di cartoncino coperto di velluto nero, e poi di raccoglierli con uno stecchino inumidito. Con questo procedimento però si può avere dei cristalli di neve una visione molto distorta perché si è portati a scegliere cristalli particolarmente grandi e simmetrici, mentre è opportuno, di quando in quando, osservare tutto quello che cade in una determinata area nell'arco di 10-20 secondi per vedere esattamente che cosa produce una nube.

Il reticolo cristallino del ghiaccio fa sí che un cristallo di neve sia un prisma esagonale, ma non garantisce che il prisma sia tabulare cioè con la base assai piú sviluppata dell'altezza a formare una lamella esagonale larga e piatta oppure sia allungato a formare un lungo cristallo aghiforme a base esagonale. Entrambi questi abiti cristallini si trovano in natura. Gli esperimenti hanno dimostrato che l'abito di un cristallo di neve è quasi interamente una funzione della temperatura. Quando la temperatura scende al di sotto del punto di congelamento, l'abito di un cristallo passa da tabulare ad aghiforme, di nuovo a tabulare e poi ancora ad aghiforme in quattro distinte fasi.

Anche la quantità di umidità dell'aria influenza la forma dei cristalli di neve. Man mano che l'aria diventa progressivamente sempre piú soprasatura, la velocità di crescita del cristallo aumenta e il suo abito viene esaltato: gli aghi diventano piú lunghi e le lamelle diventano piú grandi e piú sottili. Le lamelle inoltre diventano dendritiche ossia ornate di bei rami. Questa ramificazione dipende dal modo in cui le molecole d'acqua dell'aria si diffondono verso la lamella. In condizioni di soprasaturazione elevata gli spigoli delle lamelle esagonali sono in grado di crescere piú velocemente dei lati perché possono raccogliere molto piú efficace-

Il cristallo di neve cresce all'interno di una nube che si è formata per convezione: aria calda e umida che sale attraverso uno strato di aria fredda e asciutta. Una gocciolina di acqua si condensa (*1*) alla base della nube al livello di condensazione dell'acqua. Cresce (*2*) mentre sale in una corrente ascendente e alla fine gela (*3*) in un cristallo di ghiaccio. Le molecole di vapor d'acqua della nube si attaccano al reticolo del cristallo, dando origine ai rami della comune dendrite di neve (*4*). Il cristallo a questo punto cade (*5*) e comincia a brinare (*6*) oppure entra in collisione con goccioline d'acqua relativamente grandi. Esce fuori dalla nube e continua ad accrescersi a spese del vapore (*7*) finché inizia la sua caduta verso il suolo al di sotto del livello di condensazione del ghiaccio. La crescita del cristallo può seguire qualsiasi sequenza di eventi. Le temperature in figura sono arbitrarie.

La tensione di vapore dell'acqua liquida (*in grigio*) a qualsiasi temperatura sotto lo zero (°C) è sempre piú alta di quella del ghiaccio alla stessa temperatura (*in nero*). Ogni volta che una gocciolina d'acqua e un cristallo di ghiaccio vengono a trovarsi vicini, la gocciolina evapora e il cristallo di ghiaccio cresce. I cristalli crescono piú rapidamente a —13 °C in quanto è qui che la differenza di tensione di vapore è massima.

Le facce di un cristallo di neve (*forme bianche*) possono essere orientate soltanto in modi particolari rispetto al reticolo esagonale delle molecole di ghiaccio (*puntini*) da cui è costituito. Alcune aree della superficie del cristallo crescono piú velocemente di altre, dando origine a forme cristalline la cui simmetria esagonale esterna riflette la simmetria secondo la quale sono legate le molecole d'acqua all'interno del cristallo di neve.

mente le molecole d'acqua. Si ramificano e i rami poi si dividono in piú rami dando origine ai comuni cristalli dendritici di neve.

Se il campo di diffusione dell'umidità attorno ai cristalli di neve che crescono fosse perfettamente simmetrico, i cristalli sarebbero perfettamente simmetrici. Ma in realtà nei cristalli di neve dendritici raramente si riscontra un elevato grado di simmetria. Sembra che ciò in parte sia dovuto al modo in cui i cristalli hanno inizio e in parte al fatto che mentre crescono cadono offrendo una sola faccia al rifornimento di materiale.

Quando una gocciolina d'acqua sovraraffreddata gela, si può trasformare o in un cristallo singolo o in parecchi cristalli diversi, a seconda della dimensione della goccia e della sua temperatura quando essa si raccoglie attorno a un nucleo. Una goccia piccola con un diametro da 0,01 a 0,05 millimetri diventa un cristallo singolo se gela a una temperatura superiore ai –20 °C; diventa due o piú cristalli se gela a una temperatura molto piú bassa. Questo è provato dal fatto che i grappoli di cristalli a forma di proiettile sono molto comuni. Nel regime di temperatura piú bassa i cristalli si formano assumendo l'abito aghiforme. La gocciolina ghiacciata è composta di parecchi cristalli cosí che l'accrescimento successivo dal vapore produce un grappolo di cristalli a forma di proiettile uniti per le punte. Si assottigliano nei punti di giunzione perché lo spazio a disposizione per crescere è limitato. A volte i grappoli si rompono in cristalli singoli a forma di proiettile; a volte si raggruppano insieme formando complessi fiocchi di neve. Quando un singolo grappolo di cristalli a forma di proiettile cade in una regione con temperatura piú alta e continua a crescere, l'estremità esterna di ciascun proiettile si sviluppa in forma tabulare o dendritica. Queste particelle tridimensionali di neve sono difficili da individuare specialmente quando, durante la caduta, si raggruppano insieme.

Anche i singoli cristalli aghiformi diritti sono comuni. Quando passano attraverso zone in cui le condizioni di formazione dei cristalli sono caratterizzate da una temperatura che si aggira intorno ai –15 °C (o salgono in una corrente d'aria ascendente nel caso siano cresciuti tra i –4 e i –8 °C) a ogni estremità si forma un cristallo tabulare o dendritico. Quando i cristalli aghiformi hanno le estremità coperte da lamelle compatte, si chiamano cristalli tsuzumi dal nome di un tamburo giapponese che ha la stessa forma. Nei cristalli tsuzumi il rapporto tra la lunghezza della colonna centrale e il diametro della lamella terminale può avere qualsiasi valore. In particolare quando la colonna centrale è breve una delle lamelle o delle dendriti terminali è generalmente molto piú grande dell'altra. Presumibilmente questa caratteristica si riscontra quando le due lamelle terminali sono cosí vicine che i loro bordi competono l'un l'altro per rifornirsi di vapore.

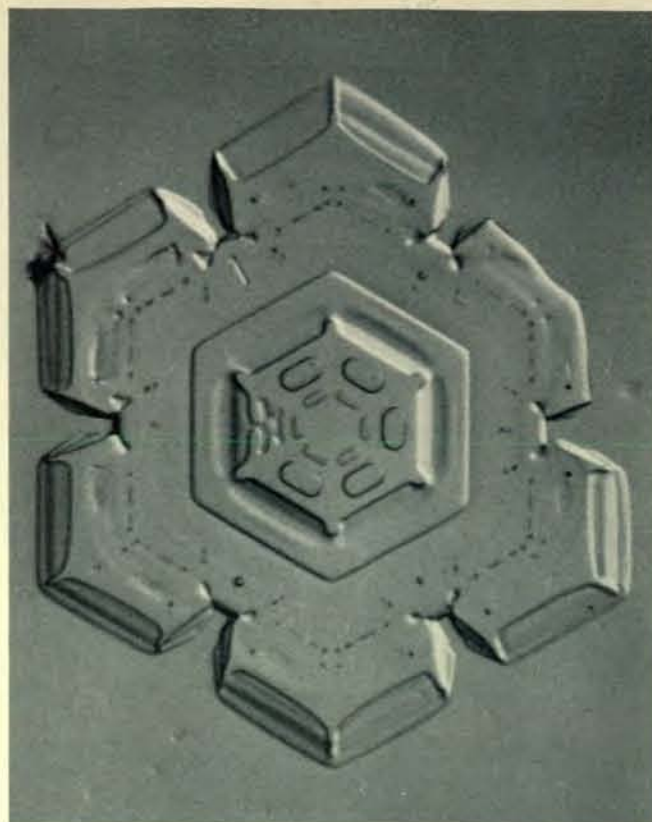
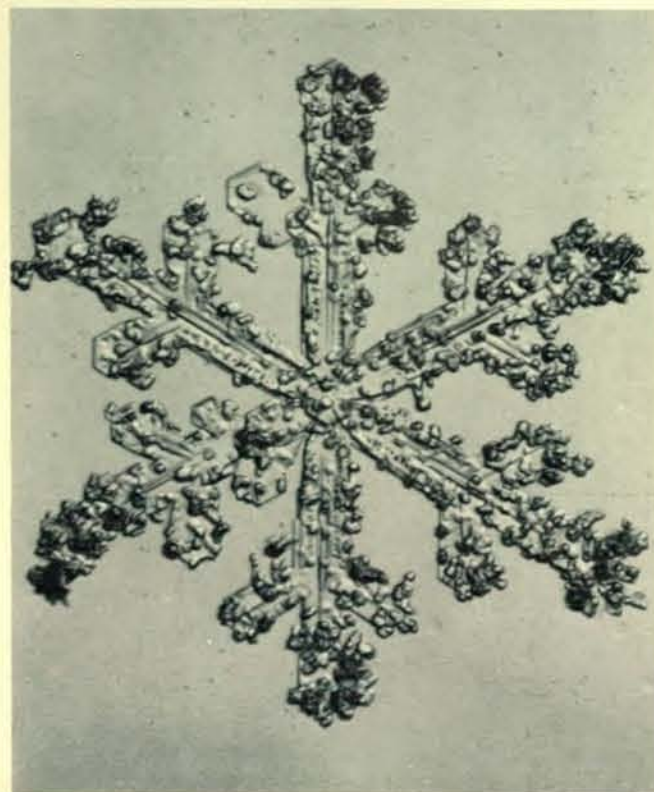
La forma di crescita di un cristallo di neve dipende dalla temperatura e dalla quantità di vapor d'acqua contenuto nell'aria. Le costituzioni si dividono in quattro regimi; in due di essi i cristalli crescono come lamelle piatte, e negli altri due come prismi e aghi. Un alto contenuto di vapor d'acqua produce costituzioni piú complesse: dendriti invece di lamelle e lunghi aghi sottili invece di aghi corti e spessi. Qualche volta i cristalli si combinano quando cadono attraverso vari regimi.

Una parte significativa – e forse perfino la maggior parte – dei comuni cristalli di neve dendritici sono in un certo senso dei veri cristalli tsuzumi, anche se hanno una « colonna » centrale estremamente corta. In realtà, la colonna centrale può essere proprio la gocciolina d'acqua originale congelatasi in un cristallo singolo con una lamella o una dendrite sviluppatasi da ciascun lato lungo l'asse di simmetria esagonale. Quando le due lamelle terminali distano meno di 0,1 millimetri, la competizione diventa molto forte e porta a cristalli di neve dendritici con meno di sei rami. In una lamella terminale si sviluppano uno, due o tre dei sei rami; nella lamella terminale opposta si sviluppano gli altri cinque, quattro o tre rami. Le due lamelle terminali possono poi separarsi, come spesso avviene, quando si posano sul velluto del raccoglitore. Ne risulta una dendrite con meno di sei rami, anche se l'angolo tra ogni coppia di rami è sempre un multiplo di 60 gradi. Quando questi cristalli non si separano, occorre un accurato e minuzioso esame per determinare che effettivamente sono composti di due strati.

Quando un cristallo di neve si sviluppa all'interno di una nube di gocce d'acqua sovraraffreddata, può crescere non soltanto sottraendo vapore intorno alle gocce ma anche collidendo effettivamente con singole gocce che immediatamente ghiacciano come accade per la brina sul terreno. Questo processo di crescita per collisione costituisce una importante complicazione in tutti i tipi di fiocchi e di cristalli di neve. Mentre un cristallo si accresce per diffusione dalle goccioline, la sua velocità, cadendo attraverso la nube, aumenta. In un primo tempo le minuscole gocce della nube sono trascinate insieme all'aria circostante il cristallo che cresce e alcune di esse evaporano completamente quando passano vicino al cristallo. Se però il cristallo diventa abbastanza grosso, le gocce piú grandi non possono né evaporare né evitare il cristallo e cosí collidono con esso.

Grappoli di cristalli a forma di proiettile si originano quando una minuscola gocciolina d'acqua gela a una temperatura molto bassa. I cristalli sono tutti uniti per le punte e l'orientamento del reticolo cristallino in ciascun « proiettile » è differente da quello di ogni altro. Se questi cristalli cadono in regimi di temperatura piú alta, all'estremità di ciascun proiettile si forma una lamella esagonale (*in alto*) o una dendrite (*in basso*).

In qualsiasi nube di un determinato tipo, un cristallo di neve mentre cade deve raggiungere una certa dimensione e una certa velocità prima di potersi accrescere per collisione in modo apprezzabile. L'accrescimento per collisione aumenta infatti la velocità di caduta del cristallo piú rapidamente di quanto non accada per l'accrescimento dal vapore. In questo caso il materiale si aggiunge soprattutto nella sua parte inferiore, aumentandone cosí il peso senza aumentarne di molto la resistenza all'aria. D'altro canto, quando il cristallo si accresce dal vapore, il materiale tende ad aggiungersi ai lati, aumentando col peso la resistenza all'aria. Questo è dunque un processo di accelerazione: piú il cristallo si accresce per collisione piú velocemente continua a collidere. Non è inusuale che i depositi formatisi in questo modo abbiano una massa molte volte superiore al cristallo originale.

I cristalli tsuzumi, chiamati cosí dal nome di un tamburo giapponese che ha la stessa forma, cominciano a crescere come una colonna centrale esagonale che poi si sviluppa a ciascuna estremità in una lamella. Qualche volta i cristalli sono chiaramente di forma tsuzumi (*a sinistra*); altre volte sono quasi simili a comuni cristalli tabulari (*a destra*). In questo caso il cristallo ha una colonna molto corta e spessa con una lamella terminale molto piú grande dell'altra. Il perfetto esagono interno è la lamella terminale piú piccola e i segni all'interno di essa sono i particolari della colonna centrale. Se una colonna centrale si sviluppa in due dendriti, una dendrite può dare origine a parte dei sei rami e la dendrite opposta ai rimanenti.

I cristalli di neve accresciutisi per collisione si sono formati catturando sia goccioline d'acqua sia molecole di vapore. Una dendrite che si è solo leggermente accresciuta per collisione (*a sinistra*) ha un particolare aspetto arrotondato. Si producono in molti casi nuovi orientamenti di cristalli quando un cristallo accresciutosi per collisione si accresce ulteriormente dal vapore (*a destra*). Il cristallo originale era una dendrite piatta; il risultato finale è noto come dendrite tridimensionale.

Un cristallo di neve può ricoprirsi di brina e poi continuare ad accrescersi dal vapore se cade nelle regioni piú basse della nube, dove le gocce d'acqua sono troppo piccole per collidere con esso, oppure se cade nella regione al di sotto della nube ma al di sopra del livello di condensazione del ghiaccio. Quando l'accrescimento per collisione produce nuovi orientamenti del reticolo cristallino, il risultato è molto pittoresco. Un tipo di cristallo noto come dendrite tridimensionale ha origine in questo modo. I nuovi cristalli non sono orientati a caso rispetto al cristallo originale, ma assumono certi orientamenti nuovi correlati a quello vecchio in modo tale che le strutture del reticolo tendono ad adattarsi l'una all'altra attraverso l'interfaccia. Probabilmente la formazione nell'accrescimento per collisione di nuovi orientamenti del cristallo è in teoria identica al congelamento delle gocce d'acqua a basse temperature in parecchi cristalli.

Quando l'accrescimento per collisione continua a tal punto che il cristallo originale non è piú riconoscibile o quasi, la neve che ne risulta è chiamata nevischio granulare. Le nubi devono essere spesse o le correnti ascendenti piuttosto forti perché i cristalli rimangano in esse abbastanza a lungo da trasformarsi in nevischio granulare. Poiché in una nevicata le particelle di nevischio granulare hanno una massa superiore e una velocità di caduta piú alta di qualsiasi altra particella, può darsi che esse rappresentino una fase di formazione della pioggia. Molte volte la pioggia è nevischio granulare sciolto. Quando le basi delle nubi sono piuttosto alte e al di sotto vi si trova uno spesso strato di aria calda e sottosatura, i singoli elementi della precipitazione che si trovano nella nube devono essere grandi per poter cadere al suolo come pioggia senza evaporare lungo il percorso. Il meccanismo che produce il nevischio granulare mette in grado molte nubi di produrre pioggia ed è questo meccanismo che i « maghi della pioggia » tentano di incoraggiare aggiungendo alle nubi nuclei artificiali di congelamento. Le particelle di nevischio granulare sono anche gli embrioni della grandine (si veda l'articolo *La grandine* di Charles e Nancy Knight, in « Le Scienze », n. 35, luglio 1971). Per impedire la formazione di grandine le nubi vengono « seminate » con nuclei che favoriscano il congelamento delle gocce d'acqua sovraraffreddata in una maggior quantità di nevischio granulare. Lo scopo è quello di far sí che i chicchi di grandine che si formano siano piú numerosi e piú piccoli e quindi si sciolgano prima di raggiungere il suolo.

Una gran parte dei fenomeni collegati ai cristalli di neve non è ancora del tutto chiara. Il problema classico è quello di spiegare perché l'abito dei cristalli varia come fa. Altri aspetti come la formazione di nuovi orientamenti nell'accrescimento per collisione e l'origine di certi tipi di nevischio granulare non sono ancora completamente conosciuti. Il problema però di gran lunga piú importante è quello sorto abbastanza di recente. Nel corso degli ultimi anni molti ricercatori hanno misurato nelle nubi che si formano nella stessa aria il rapporto tra la concentrazione dei nuclei di ghiaccio e la concentrazione dei cristalli di ghiaccio. In particolare alle temperature negative piú calde, attorno ai −10 °C, si è trovato che vi è una quantità di cristalli di ghiaccio migliaia di volte superiore a quella che ci si dovrebbe aspettare misurando la concentrazione dei nuclei di ghiaccio. Questa discrepanza ci ha portati a supporre che esista un processo secondo il quale un cristallo di neve o una particella di ghiaccio possa dare origine a un numero molto superiore di cristalli di ghiaccio. Intensi studi di laboratorio non sono riusciti finora a provare l'esistenza di questo processo. Il problema è insoluto: o le misurazioni della concentrazione di nuclei di ghiaccio sono sbagliate o il processo di moltiplicazione dei cristalli esiste e fino a ora è sfuggito semplicemente alle indagini. Una volta che tali processi siano compresi può darsi sia possibile prevederli e modificarli.

Concludiamo con una considerazione che porta al proverbiale dilemma se vi possano mai essere due cristalli di neve identici. Un cristallo di neve tipico pesa circa 10^{-6} grammi. Se la quantità media di neve che si forma sulla Terra ogni anno (compresa la neve che si scioglie o evapora prima di raggiungere il suolo) è equivalente a uno strato di acqua liquida di tre centimetri che ricopra l'intera superficie della Terra (probabilmente è una valutazione per difetto) e se la Terra ha, dal punto di vista dei fenomeni meteorologici, 3 miliardi di anni, allora in questo periodo di tempo si sono formati qualcosa come 10^{35} cristalli di neve. Questo dà 10^{29} grammi, circa 50 volte la massa della Terra. Ogni cristallo di neve, però, consiste di circa 10^{18} molecole d'acqua. Considerando l'enorme varietà di modi in cui si può disporre un cosí vasto numero di molecole, può darsi benissimo che non ci siano mai stati due cristalli identici.

Le reazioni fotonucleari

Una rassegna degli aspetti piú salienti e di maggiore attualità scientifica di questo importante capitolo della fisica e della chimica nucleari, esaminati dal punto di vista del metodo della attività indotta

di Vito di Napoli

La trasformazione di un elemento chimico in un elemento diverso da quello di partenza fu il sogno irraggiungibile degli antichi alchimisti. Questo sogno si è parzialmente realizzato nel corso degli ultimi cinquant'anni a partire dalle esperienze di Rutherford del 1919, passando attraverso le esperienze di Joliot e Curie, fino a giungere ai nostri giorni, senza far ricorso, però, a quell'apparato esoterico di cui amavano circondarsi gli alchimisti e senza bisogno di alcuna pietra filosofale. Parzialmente, si è detto, per ragioni eminentemente economiche; è possibile, in effetti, ottenere oro dal piombo, ma il costo dell'oro cosí prodotto sarebbe talmente alto da renderne irrealizzabile la produzione su scala industriale.

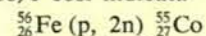
La trasformazione di un elemento in un altro è un processo che interessa il nucleo atomico. Variando il numero dei nucleoni costituenti il nucleo (con il termine « nucleone » si suole indicare uno qualunque dei costituenti il nucleo e, cioè, indifferentemente il protone p o il neutrone n) si ottiene un nuovo nucleo diverso da quello di partenza in quanto diverso è il suo numero di protoni Z o il suo numero di neutroni $N = A - Z$ (A è il numero di massa, somma del numero di protoni e del numero di neutroni) o entrambi.

D'ora in avanti, per indicare un nucleo atomico, useremo la convenzione di scrivere il simbolo chimico dell'elemento aggiungendo in alto a sinistra il numero di massa e in basso a sinistra il numero di carica; spesso si fa a meno di scrivere quest'ultimo numero, il simbolo chimico indicando già con chiarezza il numero di carica Z.

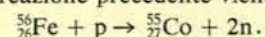
Prendiamo in esame il nucleo di cobalto $^{59}_{27}Co$ e supponiamo di riuscire, in qualche modo, a diminuire di un'unità il numero di protoni lasciando inalterato quello di neutroni; il prodotto finale di questa trasformazione è un nucleo di ferro: $^{58}_{26}Fe$.

Tali trasformazioni si ottengono con le reazioni nucleari che hanno luogo allorché si bombarda un insieme di nuclei con proiettili particolari aventi energia sufficiente per penetrare nei nuclei bersaglio e interagire con essi. I proiettili possono essere di vario tipo: elettroni, protoni, neutroni, nuclei leggeri quali $^{4}_{2}He$, $^{2}_{1}H$, o anche radiazioni elettromagnetiche di lunghezza d'onda molto piccola (raggi γ). Esistono, d'altra parte, macchine particolari – acceleratori e reattori nucleari – in grado di fornire a tali particelle l'energia necessaria per iniziare una reazione.

La maniera piú semplice per simboleggiare una reazione nucleare consiste nello scrivere per primo il nucleo bersaglio, poi, entro una parentesi e separate da una virgola, il tipo di particella incidente e le particelle fuoriuscenti dal nucleo e infine il nucleo prodotto di reazione. Per esempio, la reazione indotta da protoni, che porta da ^{56}Fe a ^{55}Co, è cosí indicata

$$^{56}_{26}Fe\,(p,\ 2n)\ ^{55}_{27}Co$$

Esiste un secondo tipo di scrittura, meno compatto del precedente, ma di piú facile lettura, che segue da vicino la convenzione adottata per le reazioni chimiche; con quest'ultimo simbolismo la reazione precedente viene scritta

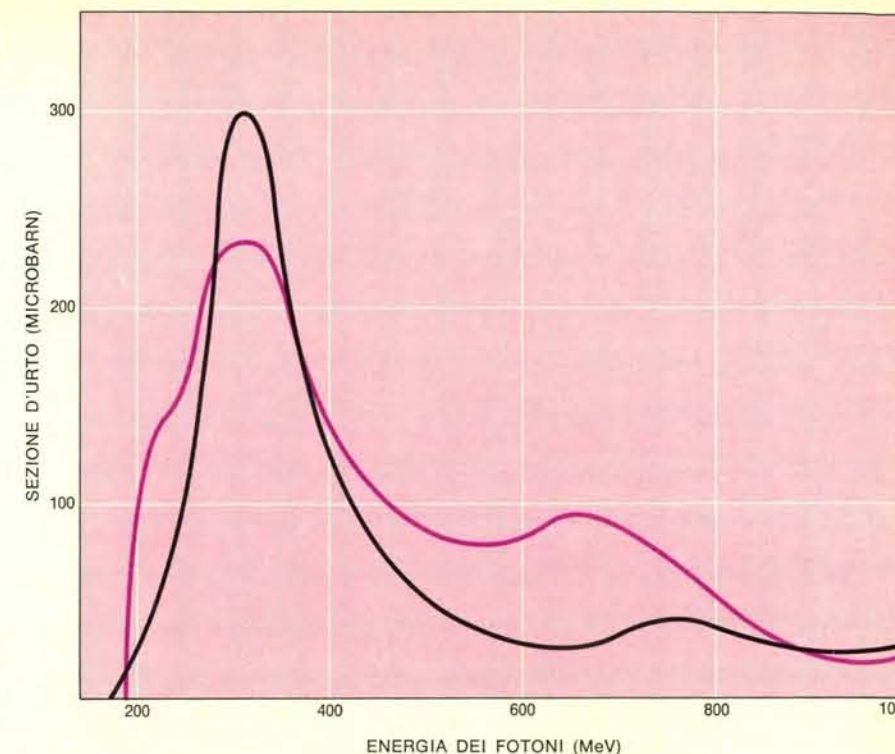
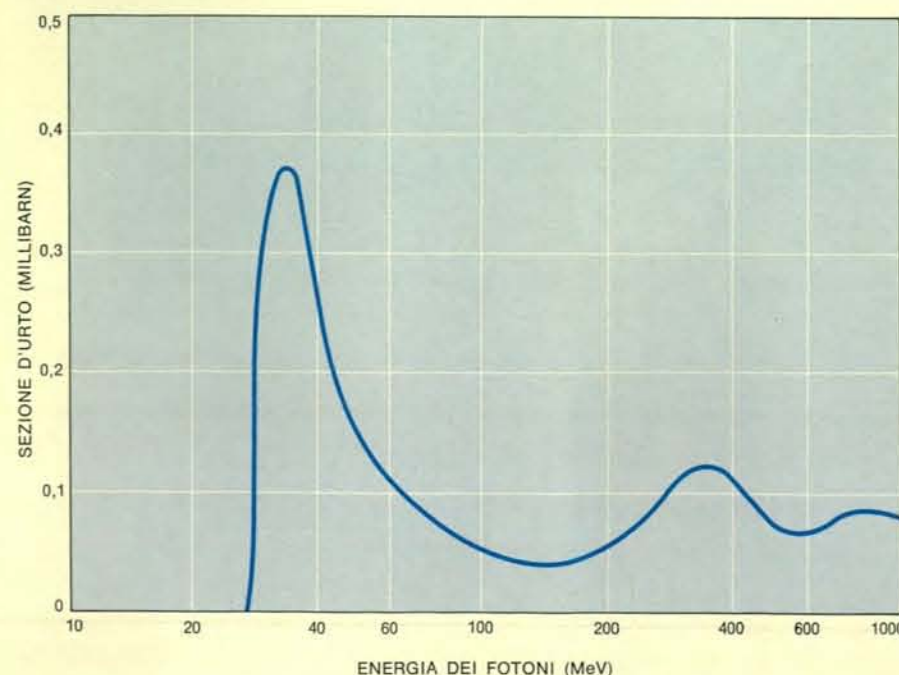
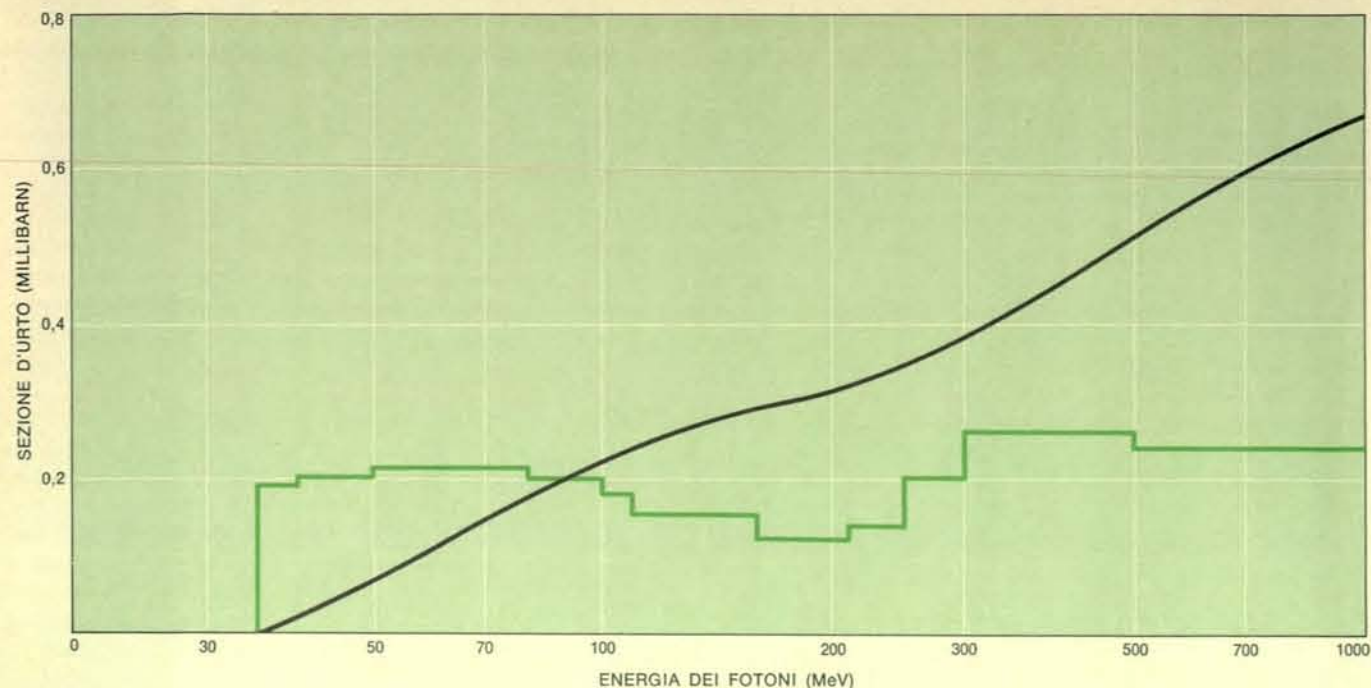
$$^{56}_{26}Fe + p \rightarrow\ ^{55}_{27}Co + 2n.$$

Andamento della sezione d'urto di fotoproduzione del berillio-7 dal carbonio-12 in funzione dell'energia dei fotoni incidenti. Sono nettamente distinguibili le tre zone che corrispondono, nell'ordine, alla zona della risonanza gigante fino a 50 MeV, alla zona del quasi deutone da 50 MeV a 150 MeV e alla zona di produzione dei fotomesoni.

Un punto di fondamentale importanza che va subito chiarito è il seguente: bombardando con protoni un insieme di nuclei di ^{56}Fe – insieme molto popolato, dell'ordine di 10^{20}, 10^{22} nuclei – solo una parte di questi interagisce in qualche maniera con i protoni incidenti; di questi, poi, una parte relativamente piccola subisce un'interazione che dà luogo a una reazione nucleare e infine, una parte ancora piú piccola dà luogo alla formazione di nuclei di ^{55}Co. Reazioni nucleari concomitanti possono essere, per esempio, $^{56}Fe\,(p,\ n)\,^{56}Co$, $^{56}Fe\,(p,\ \gamma)\,^{57}Co$, $^{56}Fe\,(p,\ p\,n)\,^{55}Fe$ e altre. Possiamo immaginare che l'evento iniziale sia $^{56}Fe + p \rightarrow\ ^{57}Co$; questo nucleo si trova in uno stato fortemente eccitato (nucleo composto) e decade, in un tempo piú o meno breve, in diversi modi e con diverse probabilità: un certo numero di tali nuclei può decadere per emissione di un neutrone dando luogo a ^{56}Co, altri decadono per emissione di quanti γ dando luogo a ^{57}Co in uno stato notevolmente meno eccitato di quello iniziale, e cosí via.

Una reazione fotonucleare, in particolare, è una reazione nucleare iniziata da fotoni, durante la quale vengono emessi dal nucleo bersaglio una o piú radiazioni γ, protoni, neutroni, aggregati di nucleoni e mesoni. Il numero di nucleoni che possono essere emessi dal nucleo durante una reazione nucleare può essere talvolta molto elevato; in alcuni casi si ha addirittura la scissione del nucleo bersaglio in due frammenti piú leggeri (fissione).

I prodotti finali delle reazioni nucleari possono essere nuclei stabili o, piú spesso, nuclei radioattivi. Dal punto di vista della chimica nucleare e della radiochimica è quest'ultimo il caso piú interessante in quanto la determinazione qualitativa e quantitativa di questi nuclei radioattivi permette di risalire alla determinazione di grandezze di fondamentale importanza per lo studio e l'interpretazione teorica delle reazioni nucleari. Su quanto esposto immediatamente sopra si basa appunto il metodo della radioattività indotta che esamineremo in dettaglio. Essendo tale metodo applicabile a qualunque tipo di reazione nucleare, nella descrizione generale di esso non si farà distinzione tra reazioni fotonucleari e reazioni indotte da altre particelle; nel seguito, invece, limiteremo il nostro discorso alle sole reazioni fotonucleari.

L'attivazione dei materiali

Se durante un certo intervallo di tempo t un insieme di nuclei bersaglio è raggiunto dal numero complessivo

In questo diagramma è mostrato l'andamento della sezione d'urto di fotoproduzione da neutroni dei pioni neutri (reazione $\gamma + n \rightarrow n + \pi^0$, *curva in nero*) e dei pioni negativi (reazione $\gamma + n \rightarrow p + \pi^-$, *curva in colore*); l'andamento della sezione d'urto di fotoproduzione è naturalmente anche qui espresso in funzione dell'energia dei fotoni incidenti.

Rappresentazione schematica dell'andamento della sezione d'urto di fotoproduzione di carbonio-11 da carbonio-12 in funzione dell'energia dei fotoni incidenti. Anche in questo caso sono nettamente distinguibili i tre picchi corrispondenti ai tre diversi intervalli di energia. La struttura reale del picco della risonanza gigante (cioè la sua rappresentazione grafica) è molto piú complessa di quella riportata per semplicità in figura.

Rappresentazione schematica della sezione d'urto di fotoproduzione di sodio-24 da alluminio-27 in funzione dell'energia dei fotoni incidenti. La curva continua rappresenta l'andamento della cosiddetta sezione d'urto per quanto equivalente, sezione d'urto introdotta per tener conto del fatto che l'energia dei fotoni di bremsstrahlung ha una distribuzione continua. La curva a gradini rappresenta invece l'andamento della cosiddetta sezione d'urto assoluta, ottenuta dalla precedente con un procedimento matematico particolare e corrispondente, intervallo per intervallo, a un ben determinato valore dell'energia dei fotoni incidenti.

N_p di proiettili e si ottiene un numero $\mathcal{N}$ di eventi che ci interessano (nell'esempio prima citato, $\mathcal{N}$ è il numero di nuclei di ^{55}Co che si ottengono bombardando con protoni nuclei di ^{56}Fe), definiamo la probabilità w che la reazione presa in considerazione avvenga come il rapporto tra eventi favorevoli e numero complessivo di tutti gli eventi possibili, in formula $w = \mathcal{N}/N_p$.

La probabilità w è legata alla quantità d di nuclei bersaglio per unità di volume e allo spessore s del materiale irradiato dalla relazione $w = \sigma d s$ in cui σ è una costante di proporzionalità che prende il nome di sezione d'urto e rappresenta la probabilità per un bersaglio ideale avente la densità di un nucleo per unità di volume e lo spessore unitario. Al prodotto $d s$ è piú opportuno sostituire una « densità » espressa in nuclei per unità di superficie e indicata con S, per cui la relazione precedente si trasforma nella relazione $w = \sigma S$.

Le due ultime equazioni mostrano come le dimensioni di σ siano quelle di una superficie; σ può quindi essere misurata in centimetri quadrati; essendo in generale molto piccoli i valori di σ, si preferisce, però, far uso di una diversa unità: il barn, che è pari a 10^{-24} cm^2, e i suoi sottomultipli.

Combinando tra loro la prima e la terza equazione si ottiene $\sigma = \mathcal{N}/S N_p$. Dal punto di vista fisico σ è la grandezza piú interessante. Lo studio della dipendenza della sezione d'urto di un particolare processo dall'energia delle particelle incidenti e dal numero di massa A del nucleo bersaglio può fornire interessanti particolari sulla natura del processo avvenuto e quindi sulla struttura del nucleo.

Dall'ultima equazione scritta si vede chiaramente come la misura di una sezione d'urto sia legata alle misure delle grandezze S, N_p e $\mathcal{N}$. La misura sperimentale di S è estremamente semplice; basta disporre di una bilancia e conoscere le dimensioni del campione sottoposto a irradiazione. Una lastrina di alluminio di forma circolare avente un diametro di 5 cm e uno spessore di 0,08 cm ha un valore di S pari a circa $4,7 \times 10^{21}$ nuclei al centimetro quadrato.

La determinazione di N_p è alquanto piú complicata, poiché richiede l'uso di apparecchiature particolari la descrizione delle quali, tuttavia, riteniamo esuli dalle finalità di questo articolo.

Piú laboriosa ancora è la misura di $\mathcal{N}$. In tale misura consiste la parte sperimentale vera e propria dello studio delle reazioni nucleari in generale e di quelle fotonucleari in particolare. Diversi sono i metodi di misura di $\mathcal{N}$; alcuni misurano gli eventi man mano che si producono, altri registrano il numero totale degli eventi che vengono successivamente contati; un altro metodo, quello della radioattività indotta, si limita alla misura della radioattività dei prodotti finali. Su questo metodo ci soffermeremo piú a lungo.

Dal valore di radioattività – espresso come numero di disintegrazioni nell'unità di tempo – si risale facilmente al numero di nuclei di una certa specie presenti nel campione irradiato al momento in cui si è effettuato il conteggio; per estrapolazione è quindi possibile ottenere il numero di nuclei presenti nel campione irradiato alla fine dell'irradiazione.

Se si ammette che il flusso di particelle incidenti si sia mantenuto costante durante l'irradiazione, indicando con Φ tale flusso, con A la superficie del bersaglio e con t la durata dell'irradiazione, si ottiene per N_p il valore $\Phi A t$ e, di conseguenza, la relazione

$$\mathcal{N} = \sigma S \Phi A t.$$

Questa equazione non tiene conto del decadimento della specie radioattiva durante l'irradiazione; prendendo in considerazione questo fattore l'equazione scritta sopra si trasforma nella relazione

$$\mathcal{N} = \sigma S \Phi A (1 - e^{-\lambda t})/\lambda.$$

in cui λ è la costante di decadimento del radionuclide in esame ed è legata al tempo di dimezzamento dalla relazione $t_{1/2} = 0,693/\lambda$.

In definitiva si ottiene per la sezione d'urto l'espressione

$$\sigma = \mathcal{N} \lambda / S A \Phi (1 - e^{-\lambda t})$$

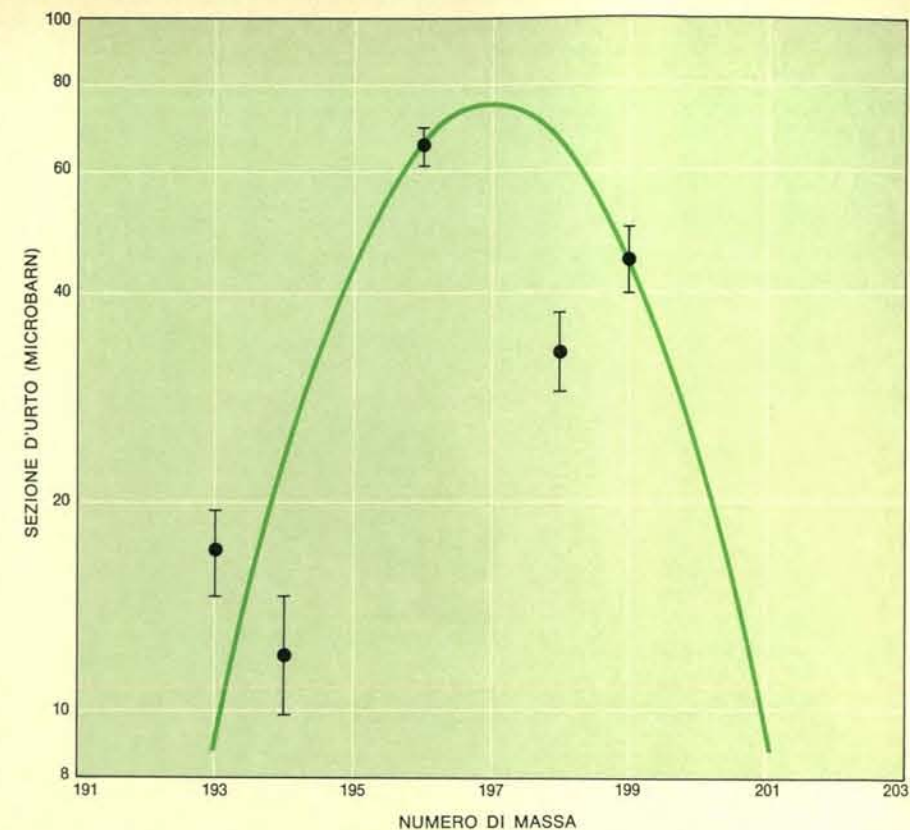
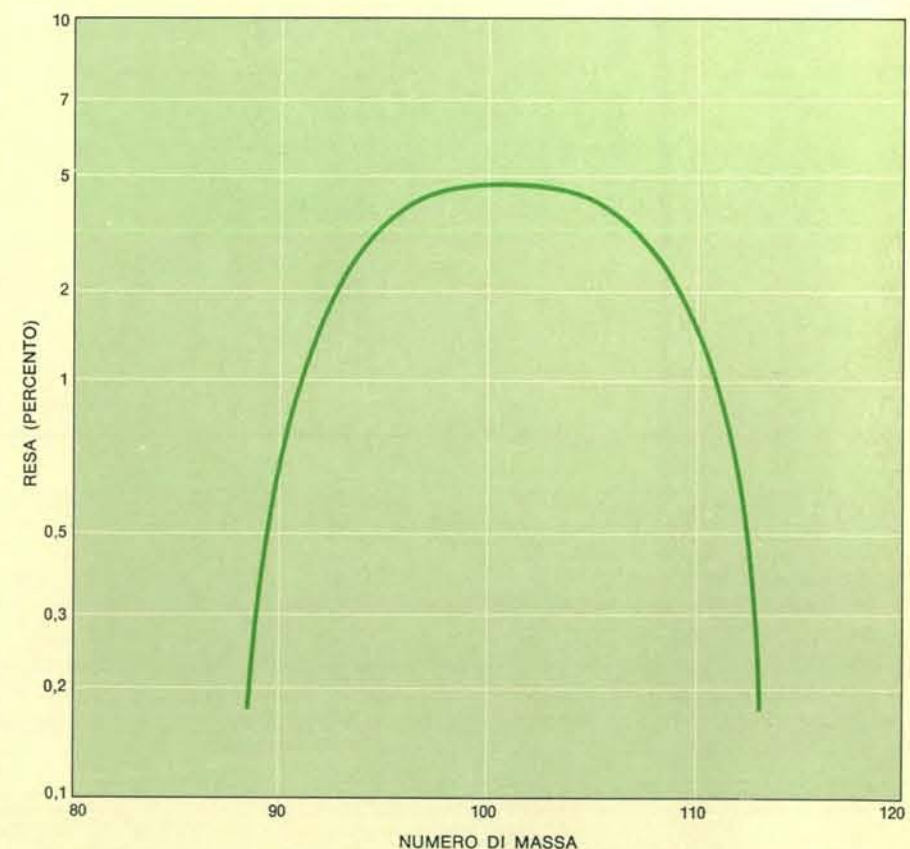
che, per piccoli valori di λt o, in altre parole, per una durata dell'irradiazione piccola rispetto al valore del tempo di dimezzamento della specie radioattiva prodotta, può essere ulteriormente semplificata sviluppando in serie l'espressione $e^{-\lambda t}$, arrestando lo sviluppo al secondo termine. Cosí facendo si perviene all'espressione $\sigma = \mathcal{N}/S \Phi A t = \mathcal{N}/S N_p$, espressione già trovata in precedenza.

Con questo metodo la misura di $\mathcal{N}$ può essere effettuata sia analizzando direttamente la targhetta irradiata, utilizzando nella quasi totalità dei casi la tecnica della spettrometria a raggi γ (cioè utilizzando le righe di raggi γ emessi dal campione), sia effettuando separazioni chimiche delle diverse specie radioattive presenti nel materiale irradiato e conteggiando le singole frazioni separate.

Come ben presto vedremo, le espressioni che danno la sezione d'urto sono di validità assolutamente generale e tali restano anche nel caso delle reazioni fotonucleari, anche se per queste ultime sorgono notevoli complicazioni a causa della natura affatto particolare della fonte di eccitazione dei nuclei bersaglio.

Sorgenti di fotoni per le reazioni fotonucleari

Fatta la dovuta eccezione per i ben pochi casi in cui è possibile ottenere fasci di fotoni monocromatici, la piú importante fonte di fotoni è la « bremsstrahlung » (radiazione di frenamento) ottenuta interponendo lungo il cammino di particelle cariche accelerate a un'energia E_o una targhetta di materiale a elevato valore di Z. L'interazione delle particelle cariche con il campo elettromagnetico dei nuclei della targhetta fa sí che tutta o parte dell'energia cinetica delle particelle incidenti si trasformi in radiazioni elettromagnetiche. I fotoni che in tal modo si originano hanno tutte le possibili energie comprese tra 0 ed E_o, secondo una distribuzione energetica di tipo approssimativamente iperbolico.

Poiché la resa in fotoni è tanto piú elevata quanto piú piccola è la massa della particella carica e quanto maggiore è il valore di Z del materiale che funge da bersaglio, nelle totalità dei casi si fa uso di elettroni accelerati e di targhette di un metallo pesante – tantalio, per esempio, o platino – i quali hanno anche un elevato punto di fusione. Gli elettroni vengono accelerati fino all'energia desiderata in macchine acceleratrici quali i sincrotroni e gli

Distribuzione di massa dei prodotti di spallazione per numero atomico pari a 79 (isotopi dell'oro) ottenuti irradiando il bismuto con radiazione di bremsstrahlung di 1000 MeV. Il massimo della sezione d'urto corrisponde all'isotopo stabile oro-197.

Distribuzione di massa dei prodotti della fissione del bismuto-209 indotta da radiazione di bremsstrahlung di 1000 MeV. Per ciascun numero di massa, la resa è il rapporto tra la sezione d'urto per quanto equivalente corrispondente a quel numero di massa e la sezione d'urto per quanto equivalente totale di fissione (10 millibarn). La massima resa si ottiene per quei valori del numero di massa che sono compresi tra 99 e 101.

acceleratori lineari e, a questo punto, vengono fatti collidere con la targhetta metallica (radiatore) ottenendo cosí un fascio di bremsstrahlung.

Il fatto che i fotoni della bremsstrahlung non siano monocromatici, a differenza di quanto avviene con altre particelle, comporta una notevole difficoltà, in sede sia teorica che pratica, nella determinazione delle sezioni di urto delle reazioni fotonucleari. Innanzitutto, al posto di N_p compare una nuova grandezza che indichiamo con Q e che prende il nome di « numero di quanti equivalenti ». Il valore di Q si ottiene dividendo per E_o l'energia totale associata al fascio di bremsstrahlung E_b.

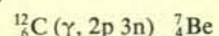
Il numero Q rappresenta formalmente il numero di fotoni che si otterrebbe se l'energia E_b fosse ugualmente distribuita tra ipotetici fotoni aventi tutti la stessa energia E_o.

Le sezioni d'urto ottenute sostituendo Q al posto di N_p o Q/t al posto di Φ nelle equazioni usate per definire σ hanno un significato diverso da quello illustrato precedentemente; nel seguito indicheremo con il simbolo σ_Q tali sezioni d'urto e le chiameremo « sezioni d'urto per quanto equivalente », distinguendole cosí dalle « sezioni d'urto assolute » σ_k (sezioni d'urto « vere », per un valore ben definito k dell'energia dei fotoni). Senza approfondire i concetti sopra esposti, ci limitiamo per ora ad asserire che è possibile, da misure di σ_Q, risalire con una certa approssimazione, piú o meno spinta, ai valori di σ_k e, nel seguito, supporremo di conoscere questa ultima grandezza che, per semplicità di scrittura, indicheremo genericamente con σ.

Dipendenza delle sezioni d'urto dalla energia. I modelli di interazione

Il metodo d'approccio piú semplice e, forse, piú valido per illustrare la variazione della sezione d'urto di una reazione fotonucleare con il variare dell'energia dei fotoni e come da tale andamento sia possibile ricavare un gran numero di informazioni utili nel formulare i diversi modelli di interazione, ci sembra quello fenomenologico.

Prendiamo in considerazione la reazione che da ^{12}C porta a ^{7}Be e che scriviamo

$$^{12}_{6}C\,(\gamma,\,2p\,3n)\;^{7}_{4}Be$$

Tale reazione ha inizio a energie dei fotoni intorno a 30 MeV (1 MeV = = 10^6 eV). A partire da questo valore di energia si ha una rapida crescita di σ fino a un valore massimo di circa 0,37 millibarn a 34 MeV; si ha poi una diminuzione abbastanza netta e continua fino a 150 MeV con un valore minimo di σ pari a 0,04 millibarn. A questo punto σ ricomincia a crescere raggiungendo un secondo massimo verso 350 MeV e un terzo massimo verso 700-800 MeV. Questo andamento è comune a tutte le reazioni fotonucleari che hanno soglie alle basse energie. (Per « soglia » intendiamo il valore minimo di energia a cui si hanno valori apprezzabili di sezione di urto.)

Possiamo suddividere l'andamento visto in tre zone distinte, a seconda dell'energia dei fotoni: una zona di bassa energia che va da qualche MeV a circa 40-50 MeV, una seconda zona di energie intermedie compresa tra circa 40 MeV e 150 MeV e una terza zona che da quest'ultimo valore si estende fino ai valori massimi attualmente disponibili (circa 20 000 MeV).

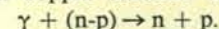
La prima zona è detta zona della « risonanza gigante » a causa dei valori relativamente elevati di sezione d'urto e dell'andamento di tipo nettamente « risonante » di σ in funzione di k. La posizione del massimo rispetto alle ascisse è caratteristica della reazione considerata, e dal numero di massa A dell'elemento bersaglio, e diminuisce al crescere di A. Il valore di sezione d'urto al massimo cresce con una certa regolarità al crescere di A fino a valori abbastanza elevati di A. In alcuni nuclei molto pesanti si ha una diminuzione di tale valore per talune reazioni e un netto aumento per altri tipi di reazione, per esempio per la fissione. Dallo studio accurato della posizione del massimo, della larghezza a mezza altezza della curva e del valore delle sezioni d'urto si è giunti alla formulazione del seguente modello di interazione. Per energie dei fotoni incidenti comprese in questo intervallo, la lunghezza d'onda del fotone è dello stesso ordine di grandezza delle dimensioni del nucleo, di conseguenza il fotone « vede » il nucleo nel suo complesso e interagisce con esso cedendogli la propria energia. Considerando il nucleo come formato da due fluidi incompressibili e non miscibili tra loro – un fluido di protoni e un fluido di neutroni –, l'eccitazione fornita dal fotone provoca lo spostamento di un fluido rispetto all'altro in maniera tale che nei pressi della superficie nucleare i due fluidi non sono piú sovrapposti. Si viene cosí a creare un dipolo sottoposto a forze di richiamo di tipo elastico proporzionali alla superficie nucleare e si origina un moto di vibrazione di dipolo la cui frequenza è caratteristica del nucleo eccitato. La diseccitazione avviene per emissione di uno o piú nucleoni.

Il massimo della sezione d'urto si ha allorché la frequenza del fotone incidente è uguale a quella della vibrazione di dipolo.

A causa della barriera coulombiana del nucleo è favorita la emissione di uno o piú neutroni rispetto a quella di protoni; in questo intervallo di energie si ha una netta predominanza di reazioni del tipo (γ, n) e (γ, 2n) anche se sono possibili reazioni diverse quali (γ, p), (γ, p n) e, nel caso di nuclei molto pesanti, (γ, fissione).

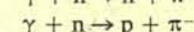
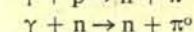
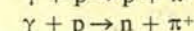
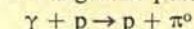
Il secondo intervallo di energia prende il nome di zona del « quasi deutone ». Il modello di interazione è il seguente. Per energie dei fotoni superiori a qualche decina di MeV, la durata della collisione tra un fotone incidente e una particella singola o piccoli raggruppamenti nel nucleo è breve rispetto alla durata delle collisioni tra nucleoni dello stesso nucleo; il primo atto dell'interazione è quindi a carico di un nucleone o di un aggregato di nucleoni.

Se la lunghezza d'onda del fotone non è ancora sufficientemente piccola da vedere un singolo nucleone, la interazione avverrà con aggregati di nucleoni, essenzialmente coppie n-p (o « quasi deutoni »); in tal caso l'evento fondamentale è la fotodisintegrazione di questa coppia secondo la reazione

$$\gamma + (n\text{-}p) \rightarrow n + p.$$

Successivamente il neutrone e il protone cosí originati possono lasciare il nucleo oppure, durante il loro cammino all'interno di questo, interagire con altri nucleoni e provocare la fuoriuscita di alcuni di questi o, ancora, cedere interamente o in parte la loro energia al nucleo il quale evaporerà un certo numero di nucleoni. La sezione d'urto di una qualunque reazione fotonucleare risulta proporzionale al numero di coppie n-p possibili nel nucleo bersaglio, numero all'incirca uguale a $8NZ/A$, e alla sezione d'urto di fotodisintegrazione del deuterio.

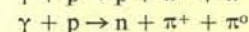
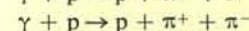
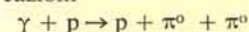
Per valori di energia compresi tra 0 e circa 150 MeV esistono regole di somma che permettono con buona approssimazione il calcolo del valore della sezione d'urto integrata, del prodotto, cioè, della sezione d'urto media totale (somma delle sezioni d'urto medie di tutti i possibili processi) in tale intervallo per il valore espresso in MeV dell'intervallo stesso.

Il terzo intervallo di energia, infine, inizia a 150 MeV e si estende fino a valori molto elevati. Al di sopra di 150 MeV la lunghezza d'onda del fotone è talmente piccola che si può supporre che l'interazione avvenga con un nucleone singolo. L'energia del fotone, inoltre, è sufficiente perché, quale primo effetto della interazione, si abbia la fotoproduzione di un mesone π (pione) secondo i seguenti possibili schemi

$$\gamma + p \rightarrow p + \pi^0$$
$$\gamma + p \rightarrow n + \pi^+$$
$$\gamma + n \rightarrow n + \pi^0$$
$$\gamma + n \rightarrow p + \pi^-$$

(*si veda la figura in alto a pagina 89*).

Dopo questo primo evento molto rapido, il mesone fotoprodotto e il nucleone di rinculo possono, durante il loro cammino, interagire con altri nucleoni secondo un processo a « cascata » della durata di circa 10^{-22} secondi e provocare l'emissione di piú particelle dal nucleo; dopo il processo di cascata il nucleo resta, in genere, in uno stato ancora eccitato e può diseccitarsi mediante un meccanismo simile a quello dell'evaporazione di un liquido. Il processo evaporativo, della durata di circa 10^{-15} secondi, ha come risultato l'ulteriore fuoriuscita di particelle dal nucleo. Competitivo con questo processo è quello di fissione. Questo modello di interazione, nelle linee generali molto simile al precedente, fa dipendere la sezione d'urto di una reazione fotonucleare dal numero di nucleoni presenti nel nucleo bersaglio e dalla sezione d'urto di fotoproduzione di mesoni π da nucleoni singoli.

Per energie molto elevate, a partire da circa 400 MeV, è possibile la fotoproduzione di una coppia di mesoni π da un singolo nucleone secondo le reazioni

$$\gamma + p \rightarrow p + \pi^0 + \pi^0$$
$$\gamma + p \rightarrow p + \pi^+ + \pi^-$$
$$\gamma + p \rightarrow n + \pi^+ + \pi^0$$

(reazioni simili avvengono a carico del neutrone) e, di conseguenza, per tali valori di energia c'è da tener conto anche di questo fenomeno. Sono altresí possibili interazioni piú complesse.

Al di sopra di 1000 MeV, comunque, l'andamento di σ resta pressoché costante o, tutt'al piú, decresce lentamente all'aumentare dell'energia.

Alcuni tipi di reazioni fotonucleari

Vogliamo ora esaminare in maggior dettaglio i tipi piú rappresentativi di reazioni fotonucleari illustrandoli con alcuni esempi.

La reazione formalmente piú semplice è quella relativa alla emissione di un solo nucleone dal nucleo, per esempio la reazione ^{12}C (γ, n) ^{11}C (*si veda la figura in basso a pagina 89*). La semplicità di scrittura, tuttavia, non trova riscontro in un'altrettanto semplice interpretazione del meccanismo di reazione per energie dei fotoni al di sopra della zona della risonanza gigante. Le difficoltà sorgono allorché si consideri che, dei prodotti di dissociazione del quasi deutone o dei prodotti dell'interazione del fotone con il nucleone singolo per energie superiori a 150 MeV, è necessario che fuoriesca il solo neutrone nel primo caso, senza, d'altra parte, che il protone fornisca al nucleo energia sufficiente per l'evaporazione di altri nucleoni; nel secondo caso è poi necessario che entrambi i prodotti (mesone e nucleone di rinculo; nel caso particolare esaminato, mesone π^0 e neutrone o mesone π^- e protone, dovendo l'interazione iniziale avvenire a carico di un neutrone) fuoriescano dal nucleo senza fornire a questo alcuna energia di eccitazione. Di conseguenza la sezione d'urto per un simile processo deve essere proporzionale alla trasparenza del nucleo nei riguardi delle particelle che debbono essere espulse e alla probabilità di assorbimento di quelle che, invece, debbono rimanere nel nucleo.

Valori di trasparenza alle diverse energie e per le diverse particelle sono stati calcolati con metodi diversi, ma, pur avendo riscontrato un ottimo accordo tra i valori sperimentali e quelli calcolati per reazioni piú complesse quali le reazioni di spallazione e di fissione che tra breve esamineremo, un meccanismo semplice come quello ora postulato non sembra essere del tutto confortato dal responso sperimentale. Molto probabilmente intervengono altri fattori quali possibili interazioni dirette o ulteriori limitazioni alla fuoriuscita delle particelle dal nucleo.

Le reazioni di fotospallazione sono caratterizzate dalla emissione di piú particelle dal nucleo, talvolta in numero assai elevato. Esempi possono essere quello già citato della fotoproduzione di ^{7}Be da ^{12}C secondo la reazione ^{12}C (γ, 2p 3n) ^{7}Be e altri quali ^{27}Al (γ, 2p n) ^{24}Na e ^{209}Bi (γ, 4p xn) $^{205-x}Au$, con x variabile tra 6 e 12 (*si veda la figura a pag. 90 e la figura in alto a pag. 91*). È interessante notare come, a parità di Z, i valori di sezione d'urto di fotoproduzione dei diversi isotopi di uno stesso elemento si dispongono approssimativamente su una parabola in un diagramma che riporti in ascissa il numero di massa A dei singoli isotopi e in ordinata il logaritmo del valore della sezione d'urto; il massimo di questa parabola corrisponde al valore di numero di massa dell'isotopo piú stabile (nel caso dell'oro, A = 197).

Interesse rilevante riveste la fotofissione. La scissione del nucleo in due frammenti di massa molto piú leggera del nucleo di partenza è, di solito, preceduta dall'emissione di particelle, in maggioranza neutroni. Dal punto di vista della chimica nucleare è molto importante ricavare alle varie energie la curva di distribuzione della resa (sezione d'urto) dei diversi frammenti di fissione in funzione del loro numero di massa. Alle basse energie la distribuzione presenta due massimi distinti, mentre alle alte energie la distribuzione è grosso modo di tipo parabolico, cioè con un solo massimo (*si veda la figura in basso a pagina 91*). I due diversi andamenti della distribuzione si spiegano col fatto che alle basse energie l'evento piú probabile è la scissione del nucleo in due frammenti di massa diversa – uno piú leggero, l'altro piú pesante – mentre alle alte energie maggiore probabilità ha la fissione simmetrica, ossia la scissione del nucleo in due frammenti di massa all'incirca uguale.

Un tipo di reazione notevolmente complesso è, infine, la fragmentazione, vale a dire l'espulsione dal nucleo di frammenti relativamente abbastanza pesanti (nuclei di Be, C, N, ecc.). Questo tipo di reazione è assai poco probabile a bassa energia, divenendo sempre piú probabile col crescere della energia. Allo stato attuale, però, ben poco si conosce sul meccanismo di interazione dei fotoni col nucleo che porta alla fragmentazione.

Conclusione

Una prima conclusione che si può trarre da quanto abbiamo esposto è la seguente: i fotoni interagiscono con il nucleo in maniera differente a seconda della loro energia; l'energia di eccitazione ceduta al nucleo, inoltre, dipende dal meccanismo di interazione e dal cammino medio delle particelle fotoprodotte nella materia nucleare.

Altre conclusioni sono, purtroppo, difficilmente proponibili allo stato attuale delle conoscenze in questo campo.

Una situazione del tutto differente si riscontra nel caso delle reazioni indotte da protoni per le quali gli studi sono attualmente di gran lunga piú estesi e approfonditi. Lo studio delle interazioni dei protoni con i nuclei complessi ha notevolmente contribuito alla conoscenza delle proprietà nucleari e del comportamento delle particelle elementari nella materia nucleare.

Poiché lo studio delle interazioni dei protoni con i nuclei, quello dei fotoni e quello degli elettroni si completano gli uni con gli altri, una sempre migliore comprensione delle proprietà dei nuclei è da attendersi col progredire degli studi nel campo delle interazioni dei fotoni e degli elettroni con la materia nucleare. Dati piú cospicui sulle reazioni fotonucleari forniranno un contributo non indifferente ad altri campi della scienza quali la fisica dei raggi cosmici e l'astrofisica.

GIOCHI MATEMATICI

di Martin Gardner

Un problema topologico con una svolta imprevista, e altri otto nuovi divertenti quesiti

Ai nove problemi che seguono, e che sono indipendenti l'uno dall'altro, risponderemo nel prossimo numero. La maggior parte di essi sono stati segnalati da lettori della presente rubrica, e appaiono qui per la prima volta.

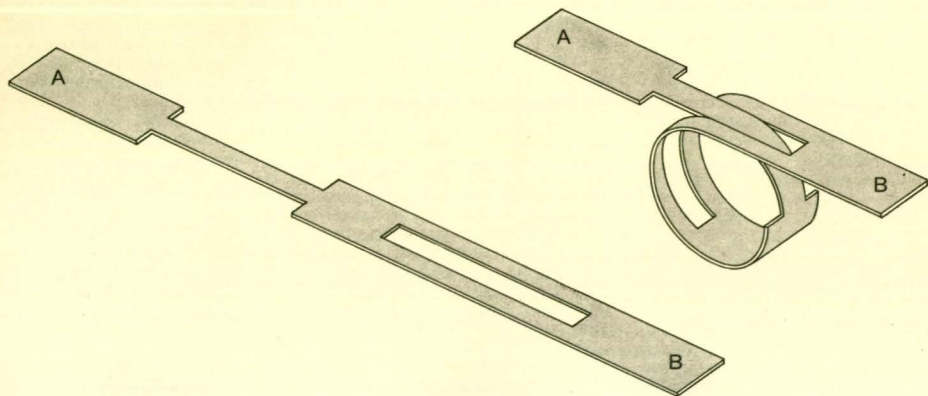
1. Gustavus J. Simmons, incaricato di ricerca e sviluppo alla Rolamite Inc. di Albuquerque, Nuovo Messico, ha inviato questo curioso problema topologico. Il lavoro che si fa alla Rolamite riguarda, tra l'altro, complessi sistemi per la laminazione di nastri metallici; uno dei tecnici della società, Virgil Erbert, si trovò un giorno di fronte al problema illustrato nella figura in basso. L'estremità *A* di un nastro flessibile era saldata a un oggetto troppo grande perché potesse passare attraverso la fessura della parte *B*. Era indispensabile, tuttavia, far assumere al nastro la forma a occhiello che si vede nella figura senza staccare l'estremità *A* dall'oggetto cui era saldata. Il problema è risolvibile?

Sembra impossibile, ma la risposta è sí. Il lettore è invitato a disegnare un facsimile del nastro su un foglio di carta, a ritagliarlo e incollarne l'estremità *A* al bordo del tavolo. Il problema, che in effetti non è difficile, consiste nel manipolare il nastro fino a fargli assumere la configurazione a occhiello.

2. Sei giocatori – che possiamo chiamare *A*, *B*, *C*, *D*, *E* e *F* – siedono attorno a un tavolo rotondo diviso in sei settori uguali. Al centro del tavolo c'è un disco, montato su un perno centrale attorno al quale è libero di ruotare (*si veda la figura in alto nella pagina a fronte*). Il disco è contrassegnato da frecce e numeri nel modo indicato dalla figura.

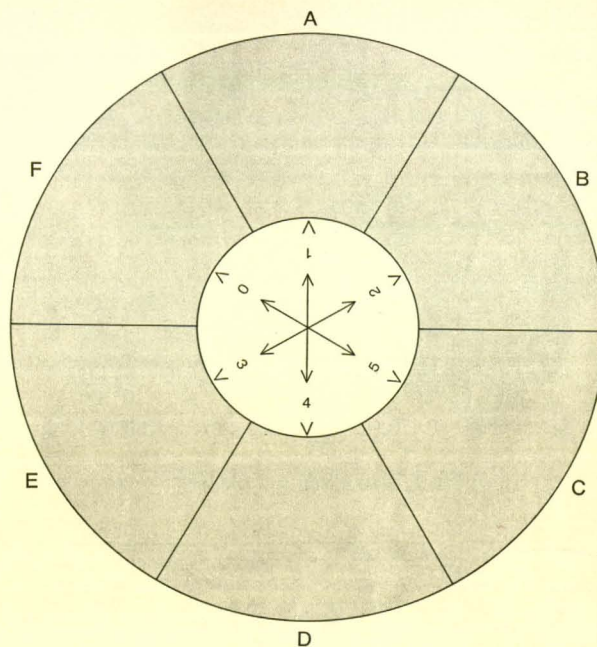
Esso vien fatto ruotare cinque volte. Dopo ogni rotazione ciascun giocatore annota il punteggio segnato all'interno della propria porzione di tavolo. (Se le frecce del disco si fermano esattamente a metà fra due giocatori adiacenti si procede a una nuova rotazione.) Vince il giocatore che realizza il massimo totale di punti dopo la quinta rotazione; nel caso il punteggio massimo sia condiviso da piú giocatori, nessuno di essi vince e l'intera gara deve essere ripetuta.

Il risultato della prima rotazione è quello mostrato dalla figura: *C* è in testa con cinque punti. Dopo il secondo giro è in testa *D*. Dopo il quinto giro la vittoria tocca ad *A*. Qual è il punteggio finale di ciascun giocatore? A prima vista sembra che i dati forniti siano insufficienti, ma il problema può essere risolto esattamente per mezzo del ragionamento deduttivo. Questo insolito quesito logico è una versione adattata di un problema pubblicato da D. Barnard nella popolare rubrica di rompicapi che quest'ultimo tiene sul giornale britannico « Observer ».

3. Un fregio è, in generale, una configurazione che si ripete all'infinito lungo una determinata striscia. Tali configurazioni possono presentare diversi tipi di simmetrie fondamentali, ma qui ci interessa soltanto quella denominata « simmetria di scorrimento ». Uno scorrimento consiste in uno slittamento (piú tecnicamente, una traslazione) combinato con una riflessione speculare e una semirotazione. Per esempio,

Il problema topologico del nastro piegato a occhiello.

Il gioco di Barnard.

facendo scorrere ripetutamente la lettera *R* verso destra lungo una determinata striscia si ottiene il seguente fregio:

RʁRʁRʁRʁRʁRʁRʁR...

H. S. M. Coxeter, professore di geometria all'Università di Toronto, ha recentemente svolto una profonda indagine su una categoria piuttosto notevole di configurazioni a fregio, che possono essere costruite molto semplicemente mediante i numeri interi non negativi, a patto di trascurare l'asimmetria dovuta alla forma dei numerali o cifre (*si veda la figura in basso in questa pagina*). Si immagini che i numerali rappresentino punti colorati: tutti gli 1 dello stesso colore, tutti i 2 di un altro colore, e cosí via. In tal caso ogni porzione rettangolare del fregio che abbia un'ampiezza di nove colonne, come quella che appare in colore nella figura, può essere considerata come la configurazione unitaria. Facendola scorrere a destra o a sinistra, e cioè facendola slittare e contemporaneamente sottoponendola a riflessione e inversione, si viene a costituire il fregio.

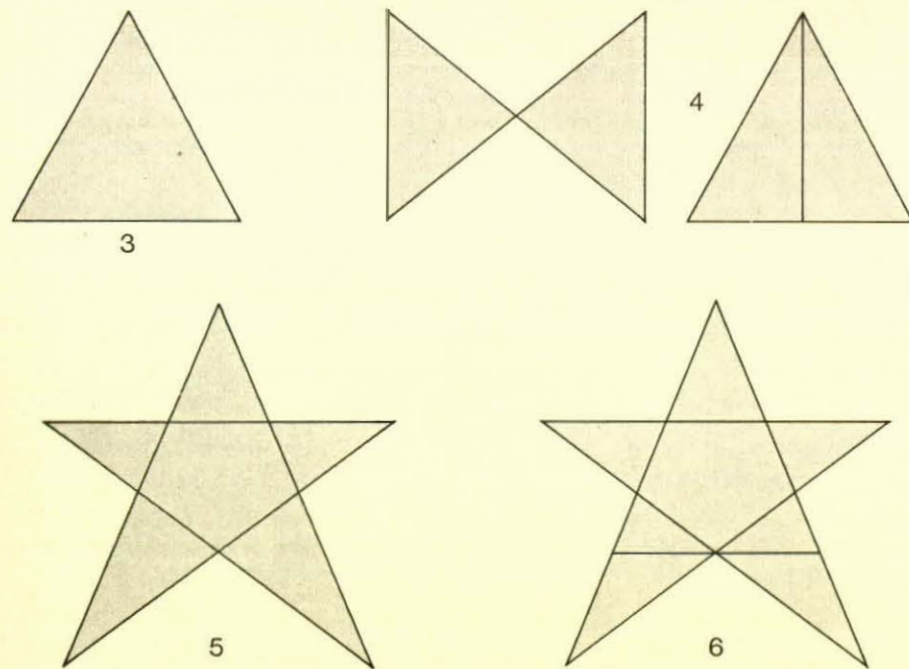
Per costruire questo tipo di configurazione a fregio si può iniziare con i due, margini inferiore e superiore, costituiti da righe infinite di 0 e di 1, e con un « sentiero » di numeri dall'alto verso il basso, per esempio come quello a zig-zag formato da otto 1 che congiunge, a sinistra nella figura, i due bordi di 0. I numeri che compaiono su questo percorso (che può essere diritto o, come nel nostro caso, curvo), come pure la lunghezza del percorso stesso, possono essere variati in modo da dar luogo a differenti configurazioni. Per ottenere tutti gli altri numeri interi si deve a questo punto applicare una semplice regola di costruzione, che è sempre la stessa per tutte le configurazioni di questo genere. La sorprendente simmetria di scorrimento risultante è un'interessante conseguenza dell'applicazione di tale regola.

Il quesito che proponiamo, su suggerimento di Coxeter, consiste nell'indovinare la semplice regola in questione. Un suggerimento: essa può essere scritta sotto forma di equazione lineare composta da tre termini, e non comporta altre operazioni che la moltiplicazione e l'addizione. Quando Coxeter mostrò per la prima volta il modello qui raffigurato al matematico Paul Erdös, quest'ultimo trovò la regola in venti secondi.

Un esame delle proprietà di tali fregi, dei loro interessanti precedenti storici e delle loro applicazioni ai determinanti, alle frazioni continue e alla geometria si trova nello scritto di Coxeter *Frieze Patterns* (*Configurazioni a fregio*) in « Acta aritmetica », vol. 18, 1971. Sui fregi in generale e sui loro sette tipi di simmetrie fondamentali si veda il moderno classico di Coxeter *Introduction to Geometry*, Wiley, 1961.

4. A un picnic, non molto tempo fa,

• • •	0		0		0		0		0		0		0		0		0	• • •
• • •		1		1		1		1		1		1		1		1		• • •
• • •			1		2		5		2		1		3		2		2	• • •
• • •				1		9		9		1		2		5		3		• • •
• • •			1		4		16		4		1		3		7		7	• • •
• • •		1		3		7		7		3		1		4		16		• • •
• • •	1		2		5		3		5		2		1		9		9	• • •
• • •		1		3		2		2		3		1		2		5		• • •
• • •			1		1		1		1		1		1		1		1	• • •
• • •		0		0		0		0		0		0		0		0		• • •

Una configurazione a fregio con simmetria di scorrimento.

Il numero massimo di triangoli non sovrapponentisi con tre, quattro, cinque e sei segmenti.

Walter van B. Roberts si trovò in mano una lattina di birra appena aperta. « Feci per posarla a terra, – egli scrive, – ma il terreno non era orizzontale, per cui pensai che fosse meglio berne un po' subito, allo scopo di abbassare il centro di gravità. Dato che la lattina è cilindrica, ovviamente il centro di gravità coincide con il suo centro quando è piena, e si abbassa man mano che il livello della birra decresce. Quando la lattina è vuota, però, il centro di gravità è situato nuovamente al centro. Deve esserci quindi un momento in cui il centro di gravità si trova nel punto piú basso possibile ».

Conoscendo il peso della lattina vuota e il peso della lattina piena, come si può determinare qual è il livello di birra che, in una lattina diritta, porta il centro di gravità al punto piú basso possibile? Quando Roberts e i suoi amici si misero ad affrontare il problema tirarono in ballo, a un certo punto, la analisi matematica: esprimendo l'altezza del centro di gravità in funzione del livello di birra, differenziando, eguagliando a zero e risolvendo l'equazione per il valore minimo dell'altezza del centro di gravità, il problema è risolto. Piú tardi, tuttavia, a Roberts venne in mente un semplice metodo per arrivare alla soluzione senza analisi matematica; in effetti, il problema può essere risolto semplicemente pensandoci un po'.

Per inquadrarlo in termini precisi, supponiamo che la lattina vuota pesi 15 grammi, che sia perfettamente cilindrica e che le asimmetrie dovute ai fori in essa praticati siano trascurabili. La lattina contiene 120 grammi di birra, per cui il peso totale della lattina piena è di 135 grammi; l'altezza della lattina è di 16 centimetri. Senza servirsi dell'analisi matematica, si determini il livello della birra in corrispondenza del quale il centro di gravità si trova alla minima altezza.

5. Su un tavolo ci sono tre monete: un quarto di dollaro, un mezzo dollaro e un dollaro d'argento. Smith prende una moneta e Jones prende le altre due. Tutte e tre le monete vengono lanciate in aria contemporaneamente.

Si conviene che, se una moneta dà « croce », il punteggio è zero; se dà « testa », il punteggio corrisponde al suo valore in centesimi. Chi ottiene il punteggio piú alto vince tutte e tre le monete, ma se tutte e tre le monete danno « croce » il lancio deve essere ripetuto.

Domanda: quale moneta deve scegliere per sé Smith affinché la gara sia equa, cioè affinché il guadagno monetario probabile sia zero per ambedue i giocatori?

Autore di questo nuovo ed inedito quesito è David L. Silverman, che ha pubblicato l'interessante volume di quiz intitolato *Your move* (McGraw-Hill, 1971). La risposta è sorprendente. Ancora piú sorprendente è una formulazione generalizzata di questo problema, la cui dimostrazione formale è dovuta a Benjamin L. Schwartz.

6. Kobon Fujimura, esperto giapponese di rompicapo, ha recentemente ideato un quesito di geometria combinatoria. Si tratta di un problema semplice da definire, ma del quale non si conosce ancora una soluzione generale. Qual è il massimo numero di triangoli non sovrapponentisi che si possono costruire con n segmenti di retta?

Non è difficile scoprire per tentativi che per $n = 3, 4, 5$ e 6 il massimo numero di triangoli è rispettivamente uno, due, cinque e sette (*si veda la figura a fianco*). Con sette segmenti il problema comincia a diventare difficile. Il lettore è invitato a cercare il massimo numero di triangoli non sovrapponentisi costruibili con sette, otto e nove segmenti.

Il problema di trovare una formula capace di determinare tale numero in funzione del numero di segmenti assegnati sembra essere estremamente difficile.

7. Una delle soddisfazioni che si possono trarre dai giochi matematici sta nel trovare soluzioni migliori a problemi che si riteneva fossero già stati risolti nel miglior modo possibile. Prendiamo in considerazione il seguente problema numerico, che appare sotto il n. 81 nel volume di Henry Ernest Dudeney *Amusements in mathematics*, ristampato in epoca recente dalla Dover dopo la prima edizione del 1917. Nove cifre (zero escluso) vengono suddivise in due gruppi; in quello di sinistra, un numero di tre cifre viene moltiplicato per uno di due cifre, mentre in quello di destra ambedue i numeri sono di due cifre:

158	79
23	46
—	—

In tutti e due i casi il prodotto è lo stesso: 3634. Dudeney si chiese come si possono disporre le stesse nove cifre secondo il medesimo modello in modo da ottenere un prodotto (identico a destra e a sinistra) che sia il massimo possibile. La risposta di Dudeney che, come disse egli stesso, « si può trovare solo con un po' di ragionamento e di pazienza », fu

174	96
32	58
5568	5568

In seguito, Victor Meally di Dublino migliorò notevolmente la soluzione di Dudeney, fornendo la seguente risposta:

584	96
12	73
7008	7008

Quest'ultima cifra ha segnato il record fino all'anno scorso, quando un giapponese amico di Fujimura scoprí una soluzione ancora migliore, che, sebbene ancora non vi siano prove, si ritiene sia il massimo numero possibile. È il lettore in grado di scoprire tale soluzione senza ricorrere a un elaboratore elettronico?

8. Un noto problema in materia di dama. Si dispongono otto pedine in fila; la mossa consiste nel prendere una pedina, farle scavalcare verso destra o verso sinistra esattamente due pedine e piazzarla su un'altra pedina formando una dama. Scavalvare una dama conta come scavalcare due pedine. Problema: formare quattro dame in quattro mosse. Non è un problema difficile, e inoltre si può facilmente dimostrare che per ogni numero pari di pedine n, quando n è almeno 8, è sempre possibile formare una fila di $n/2$ dame in $n/2$ mosse. Numerose varianti di questo vecchio problema sono state proposte da Dudeney e da altri autori. La seguente variazione sul tema, che ritengo sia nuova, è stata suggerita e risolta da W. Lloyd Milligan.

Si dispongono in fila n pedine, in numero pari. Prima si salta una sola pedina, formando una dama; poi se ne saltano due, poi tre, e cosí via, aumentano ogni volta di una unità il numero di pedine da scavalcare per formare una dama. L'obiettivo è quello di formare $n/2$ dame in $n/2$ mosse.

Al lettore il compito di dimostrare che il problema è risolvibile solo se n è multiplo di 4. Si possono trovare facili soluzioni per tentativi nel caso di $n = 4$ o 8, ma per $n = 16$ o per valori ancora piú alti le difficoltà diventano notevoli in mancanza di un metodo sistematico.

9. Le vignette senza parole si basano spesso su situazioni fisicamente impossibili, o che rappresentano un'offesa alla logica. Lewis Carroll amava raccontare di un tale che aveva piedi cosí grandi da doversi infilare i pantaloni dalla parte della testa. Pressappoco lo stesso genere di assurdità sta alla base di una famosa vignetta, pubblicata da Charles Addams sul « New Yorker », in cui si vede una sciatrice che fila giú per una discesa; dietro di lei, le tracce parallele lasciate dai suoi sci vanno in direzione di un grosso albero, passano attorno ai due lati del tronco e subito dopo ridiventano parallele.

Supponiamo che vi imbattiate in un paio di tracce del genere su un pendio nevoso, e che le vediate passare ai lati di un albero esattamente come nella vignetta di Addams. Diamo per scontato che siano effettivamente tracce di sci: riuscite a immaginare almeno sei spiegazioni che siano fisicamente possibili?